Are there characteristics associated with higher performing breast units that could inform recommendations to standardise arbitration processes?

Hackney, Lisa

Award date:
2021

Awarding institution:
Coventry University

Are there Characteristics Associated with Higher Performing Breast Units that Could Inform Recommendations to Standardise Arbitration Processes?

By

Lisa Hackney

September 2020

A thesis submitted in partial fulfilment of the University's

requirements for the Degree of Doctor of Philosophy

Certificate of Ethical Approval

Applicant:

Lisa Hackney

Project Title:

Are there characteristics associated with high performing breast units that could inform recommendations to standardise arbitration or consensus processes?

This is to certify that the above named applicant has completed the Coventry University Ethical Approval process and their project has been confirmed and approved as Medium Risk

Date of approval:

06 February 2017

Project Reference Number:

P45921

Certificate of Ethical Approval

Applicant:

Lisa Hackney

Project Title:

Are there characteristics associated with high performing breast units that could inform recommendations to standardise arbitration or consensus processes?

This is to certify that the above named applicant has completed the Coventry University Ethical Approval process and their project has been confirmed and approved as Medium Risk

Date of approval:

26 May 2017

Project Reference Number:

P50587

Abstract

Rationale

In recent years the overall benefit of breast screening has been a subject of debate, and one of the criticisms is excessive false-positive recalls. Arbitration by a third reader or group consensus can be integral in reducing these. Before the publication of the Public Health England arbitration guidance (August 2016), the third person arbitrator or lead of consensus meetings had to be medically qualified (radiologists, breast clinicians). How sensitive and specific the third reader should be, has never been specified, but there is considerable variation between individuals undertaking the task.

This research aimed to explore the different reporting and arbitration strategies in breast screening within England to ascertain if specific systems work better in differing units, and thereby inform recommendations to standardise processes. Consideration is also given to advances in technology in this field.

Method

A mixed-methods approach was used to explore the complex factors associated with decision making (reporting and arbitration) in breast screening and the effect on recall rates. The research included two national surveys, analysis of chosen performance metrics (recall rates, small cancer detection rates and Standardised Detection Ratio) for all 80 breast screening units in England (KC62 data) and semi-structured telephone interviews, based on a pre-determined sampling frame. Interviews were undertaken to explore the opinions, experiences, perspectives and

insights of reporting staff (varying professional roles). Methodological triangulation was used to evaluate complementary and divergent findings.

Key findings

The survey results demonstrated variability in all aspects of reporting and arbitration practices. The reporters may be influenced by non-blind reading and arbitration, resulting in biased decision-making. The PHE guidance on arbitration has had minimal impact on the respondent units.

Analysis of the KC62 data demonstrated variations in the performance parameters reviewed at the unit level, but in particular, recall rates. However, there was no difference in mean recall rates between units for the cases reviewed; the arbitration strategy; the reading type; professional role undertaking the third reader arbitration/leading consensus or programme size. Also, there were no statistically significant differences for the four-year average prevalent and incident SDR between programme sizes nor between the arbitration strategies for small cancer detection rates (prevalent and incident) or SDR (prevalent and incident).

The interview results generated five main themes relating to reporting and arbitration practices: organisational factors, technology, clinician factors, teamwork factors and PHE guidance factors.

Artificial Intelligence (AI) could potentially tackle some of the current challenges in breast screening, including capacity issues/workforce planning, increased efficiency, improved accuracy and advanced detection of early cancers. Further research is needed on optimising human/AI decision-making.

Conclusion and further research

This thesis has resulted in several organisational and national recommendations regarding blind reading/arbitration to provide improved film reader data profiles and standardisation, and considerations surrounding alternative models of service delivery.

The research has revealed the potential for future work into:

1. the design of the breast screening reporting system
2. selection of arbitrators and alternative methods of group decision making, and
3. determining cultural and organisational characteristics that may improve diagnosis and support effective teamwork

Acknowledgements

The completion of the thesis was made possible by a Coventry University funded studentship.

There are many to whom I owe acknowledgement and thanks.

I am very grateful to my project supervisors for their experience, support, guidance and constructive feedback offered throughout this PhD: Professor Derek Renshaw, Professor Ala Szczepura, Professor Louise Moody and Becky Whiteman.

I want to thank the staff who participated in this research. Particular gratitude is owed to those who participated in the interviews. They gave their time freely and were candid in giving their views as part of this work.

More personally, special gratitude to my incredible family and friends for their help, support and continuous encouragement. In particular, my partner Darren for his unwavering belief in me to achieve my ‘goals'.

Contents

Tables and Figures

Glossary

Arbitration	The use of a third reader to decide on case management when there is disagreement between the initial reporters.
Blind reading	The second reader is unaware of the first readers report.
Cancer Detection Rate	The proportion of screened women with breast cancer who test positive for breast cancer. Often expressed as a percentage.
Code	A descriptive or conceptual label that is assigned to extracts of raw data in a process called 'coding'.
Consensus	A group of film readers who decide on case management when there is disagreement between the initial reporters.
Coverage	Defined as the percentage of women in the population who are eligible for screening at a particular point in time who have had a test with a recorded result at least once within the screening round (past 36 months).
Discrepant Cancer Rate	The number which the first reader recommended be returned to routine recall which were ultimately recalled for assessment and diagnosed with cancer (shown as a rate per 1,000 women).
Double-reading	A breast screening protocol in which two film readers independently report the same images.
Eligible screening population	Women between the ages of 50 to 70 are eligible for screening who are registered with a GP. Women aged over 70 are eligible to be screened if they self-refer.
False negative	A decision made in error that a case is negative

	for cancer when the case is cancer.
False positive	A decision made in error that a case is positive for cancer when the case is cancer free.
Incident screen	Screening of women previously screened within the NHS breast screening programme who have been screened within the last 5 years.
Interval cancer	A cancer that presents clinically between screening rounds.
Invasive cancer	A malignant tumour which has spread to invade cells beyond the cell wall.
KC62	A National statistical mandatory return that all breast screening units in England are required to undertake on an annual basis.
Non-blind reading report	The second reader is aware of the first readers
Non-invasive cancer	An early form of carcinoma. There are cancerous cells, but they have not started to grow outside of the cell wall.
Positive Predictive Value	The probability of screened women with a positive (malignant) test that have breast cancer. Often expressed as a percentage.
Prevalent screen	Screening of women never previously screened within the NHS breast screening programme. Within the standards it relates to women's first ever screening appointment.
Reader	An individual trained to report breast-screening mammograms.
Recall rate	The proportion of screened women that are asked to return for further assessment. Often expressed as a percentage.
Report	The reader's final opinion on a screening

	mammogram.
Screening round length	The screening round length for the breast screening programme is 36 months and all eligible women should receive a screening invitation within 36 months of a previous screen
Sensitivity	The ability to correctly detect disease in the eligible screening population who have the disease.
Specificity	The ability to correctly exclude disease in the eligible screening population who do not have the disease.
Standardised Detection Ratio	This is the ratio of the observed number of invasive cancers to the expected number based on applying criteria from the Swedish Two Counties randomised control trial which is used as the comparator for performance. An SDR of 1 equates to parity with this trial.
Theme	An interpretive concept describing or explaining aspects of the data, following analysis of the whole dataset.
Transcript	A written verbatim (word-for-word) account of an interview.

Abbreviations

Acronym	Definition
2D	2-Dimensional
ACR	American College of Radiology
AGREE	Appraisal of Guidelines, REsearch and Evaluation
AHP	Allied Health Professional
AI	Artificial Intelligence
ANN	Artificial Neural Networks
AUC	Area Under the receiver operating characteristic Curve
BIRADS	Breast Imaging Reporting and Data System
BOS	Bristol Online Survey
BSIS	Breast Screening Information System
CAD	Computer Aided Detection
CAT	Computerised Adaptive Testing
CDR	Cancer Detection Rate
CESM	Contrast Enhanced Spectral Mammography
CI	Collective Intelligence
CME	Continuing Medical Education
CPD	Continuing Professional Development

CPG	Clinical Practice Guidelines
DBT	Digital Breast Tomosynthesis
DCIS	Ductal Carcinoma In Situ
DFS	Disease-Free Survival
DM	Digital Mammography
DMIST	Digital Mammographic Imaging Screening Trial
EBP	Evidence-Based Practice
FFDM	Full-Field Digital Mammography
FRQA	Film Reader Quality Assurance
GCP	Good Clinical Practice
GRAMMS	Good Reporting of A Mixed Methods Study
HRA	Health Research Authority
HRL	High Risk Lesion
IIQM	International Institute for Qualitative Methodology
IRAS	Integrated Research Application System
IT	Information Technology
LCIS	Lobular Carcinoma In Situ
LORIS	LOw RISk DCIS trial
MD	Mammographic Density

MDT	Multidisciplinary Team
MMR	Mixed Methods Research
MRI	Magnetic Resonance Imaging
NBSS	National Breast Screening Service
NDROR	Non-Discordant Radiographer Only Reporting
NHS	National Health Service
NHSBSP	National Health Service Breast Screening Programme
ODR	Office for Data Release
OS	Overall Survival
OTST	Oslo Tomosynthesis Screening Trial
PACS	Picture Archiving and Communication System
PERFORMS	Personal Performance in Mammographic Screening
PHE	Public Health England
PPV	Positive Predictive Value
QA	Quality Assurance
RAC	Research Advisory Committee
RCR	Royal College of Radiologists
ROC	Receiver Operating Characteristic
SCoR	Society and College of Radiographers

SD	Standard Deviation
SDR	Standardised Detection Ratio
SFM	Screen-Film Mammography
SOP	Standard Operating Procedure
SPSS	Statistical Package for the Social Sciences
TA	Thematic Analysis
TNM	Tumour–Node–Metastasis
TOMMY	TOMosynthesis with digital MammographY
URL	Uniform Resource Locator
WHO	World Health Organization

Chapter 1. Introduction

Breast cancer is a significant health burden worldwide, and the second most common cause of cancer-related deaths in UK females (Cancer Research UK 2020). Certain factors predispose an individual to a higher risk of developing breast cancer, and these can be categorised as those which are modifiable and non-modifiable. Although breast cancer incidence rates have risen or stabilised in some countries over the last decade, mortality rates have declined (Cancer Research UK 2020). Reductions in mortality are attributed to earlier diagnosis via screening and improved treatment; although the respective influences of each are uncertain (Malvezzi et al. 2019). The intent of breast screening to reduce mortality from the disease is only successful if specific measures are met, and breast screening units are monitored to ensure programme safety and effectiveness (PHE 2017).

The current technique for population-based screening is mammography, but this procedure has inherent limitations. The variability in the performance of a screening unit can be attributable to many factors relating to the characteristics of the population screened, the reporting personnel, variances in practice (recall standards, screening interval, number of readers, arbitration processes) and imaging technologies (Mohd Norsuddin et al. 2015). To increase the cancer detection rate double reading has become the standard practice in many countries (Perry et al. 2007). However, there is an international variance in how this is undertaken. Double reading inherently creates a probability that the two reporters may disagree on their radiological opinion (Klompenhouwer et al. 2015b). Discordant findings require resolution, and this is commonly achieved by some form of group consensus

or third reader arbitration. A systematic scoping review of the evidence on the use of these processes within breast screening (Hackney et al. 2017) found a limited body of evidence, and specifically a lack of prospective studies to determine effectiveness in real-life clinical settings. Only a few studies reported true interval cancer rates and many reported results with an insufficient follow-up which compromised the ability to conclude the effectiveness of the processes. Within England, there is a wide variation in recall rates with some units not achieving the NHS Breast Screening Programme (NHSBSP) standards for prevalent assessment recall (NHSBSP Central Return Data Set KC62). Arbitration can be integral to achieving this.

1.1 Clinical Resources/Skills Mix

A potential obstacle for breast screening units in sustaining the current quality standards is the chronic shortage and predicted workforce retirement of specialist Radiologists in England (The Royal College of Radiologists 2020). Concerns about the future availability of breast Radiologists are highlighted in the report, which predicts that 26% (n=134) of breast Radiologists will retire in the next five years, combined *"with a potential 2.2 million increase in women eligible for screening if the age extension is implemented (based on current population figures)"*(Moser et al. 2011). Coupled with the unification of high-risk family history screening (in particular breast Magnetic Resonance Imaging) into the NHSBSP, increasing interventional radiology procedures (diagnosis and excision of lesions by vacuum-assisted biopsy) and the potential expanded use of emergent technologies (Digital Breast Tomosynthesis) which are more labour intensive, the demand on radiology services is raised further.

1.2 Task Shifting/Role Extension

One eminent approach to addressing human resource problems is the extension of duties from medics to Allied Health Professionals (AHP's). In the literature, this was also referred to as up-skilling, role extension, task-shifting or task optimisation (Debono et al. 2015, Torres-Mejía et al. 2015, Singh et al. 2017 and Moran and Warren-Forward 2016). Task shifting is considered one method of restructuring roles and responsibilities (changing professional boundaries) to make the most effective and efficient use of skill mix (Singh et al. 2017). Radiographers in the UK have long been familiar with the concept of role extension and its associated opportunities and challenges. Wells and Cooke (1996) first report Radiographers formally undertaking mammography reporting in NHS screening units. Over the next sixteen years, this progressed to double Radiographer reporting following the Non-Discordant Radiographer Only Reporting (NDROR) trial (Bennett et al. 2012).

1.3 Public Health England Arbitration Guidance

Prior to August 2016 single third-person arbitration or lead of consensus meetings, was a responsibility only of medically qualified professions (Radiologists, Breast Clinicians). Severe breast Radiologist shortages necessitated a review of national guidance in order to maintain current quality standards and avoid delays in patient management. The revised Public Health England guidance (Public Health England 2016) (Appendix 1) recognised that the skills to perform arbitration/lead consensus were not necessarily associated with the profession of the arbitrator, and delegation of these duties would help to decrease the current pressure on services.

A survey undertaken by Culpan (2016) reported that 23% (n=15/66) of UK

Radiographers were already undertaking third reader arbitration or providing the definitive vote in discordant screening cases. This suggests that some breast services had already started to implement changes in practice through local governance systems in advance of the guidance. Individual healthcare organisations are renowned for developing distinctive professional cultures based on an evolution of local practice over time. The concepts of organisational culture and climate have been extensively described in the literature (Ginsburg and Gilin Oore 2016, Erasmus et al. 2017, Everest, Fitzgerald, and Tate 2014). Zohar and Hofmann (2012) affirm that climate relates to an employees' perceptions of procedures, practices and behaviours. In contrast, culture is entrenched and can be characterised as shared underlying assumptions, values and beliefs that typify a setting and help to explain why things happen in a particular way (Ostroff, Kinicki, and Muhammad 2013, Schneider, Ehrhart, and Macey 2013). Zohar and Hofmann (2012) emphasise that climate and culture are multilevel constructs, and employees will develop perceptions of both the organisational and sub-group climate within which they work. It was thus predicted that the philosophies regarding arbitration practices and delegation of these duties to Radiographers would be multifaceted, with a variation in the current landscape.

Given the PHE guidance (Public Health England 2016) supports delegation of arbitration duties to Radiographers, there is a need to establish the current national practice for recalls within breast screening services and to identify if this process may be improved. Evidence from the pilot data in the NDROR study (Bennett et al. 2012) suggested that double Radiographer reporting could increase recall rates but

was considered unlikely to have significant impacts on performance in the NHSBSP if

"*fully supported and carefully monitored* (particularly recall rates) (Bennett et al. 2012: 120)".

Currently, it is unknown what the implications are within England of the new PHE arbitration guidance and the potential barriers and facilitators to implementation. The researcher hypothesises that Radiographer arbitration may only be implemented in services with a severe shortage of Radiologists and not delegated in services where the radiology workforce is sufficient.

Unlike the NHSBSP standards of quality that define the acceptable and achievable levels of performance that must be adhered to, Eddy (1990) states that guidance should allow for some degree of flexibility. Three approaches are described concerning the introduction of clinical guidance: diffusion, dissemination, and implementation (Lomas 1993). Culleton (2015: 444) expresses that conventionally,

"clinical practice guidelines were consensus-based statements derived from expert opinion.

It is now accepted that they should be constructed via a transparent process to minimise bias, ensure a systematic approach to evidence collation and evaluation, with an emphasis on patient-relevant outcomes. The principles defined by Culleton (2015) are summarised in Table 1. Culleton (2015) states that CPG's frequently arise in response to a service need such as task shifting. A principal barrier identified with the adoption and use of clinical guidelines in health care relates to the provider's

overall attitude towards the guidance or uncertainty regarding its reliability.

Table 1 Proposed Principles for Clinical Practice Guidelines (CPG) Development (Taken from Culleton 2015)

• Processes for developing and evaluating CPG should focus on outcomes.
• CPG should be based on the best available evidence and graded according to the level, quality, relevance, and strength of evidence
• CPG development should be multidisciplinary and include consumers
• CPG should be flexible and adaptable to local conditions. They should include evidence for different target populations and take into account patient preferences.
• COG should be developed with resource constraints in mind.
• Implementation plans should be developed along with CPG.
• The implementation of CPG should be evaluated
• CPG should be revised regularly to account for new evidence

1.4 Thesis Aims

This research aims to explore the current variation in reporting and arbitration strategies within breast screening services in England. It seeks to correlate findings with performance based on specific criteria from published national service data (KC62 2013/2014 -2016/2017) to ascertain if there are characteristics associated with decision-making in higher and lower performance units that could inform the future effective use of existing arbitration processes. A further aim was to comprehend the future role of new technology, in particular, Artificial Intelligence (AI) in this setting.

The main objectives are:

Objective 1: Extend the systematic scoping review of published literature to identify research evidence on the barriers and facilitators for decision-making in mammogram reading and, use narrative synthesis to develop a conceptual framework.

Objective 2: Undertake two surveys of Breast Screening Units within England to explore how and why arbitration systems were established, why practice varies and the consequences of such variation for imaging professionals, and service user outcomes.

Objective 3: Collect and analyse published data (KC62) on a series of unit characteristics – the size of the unit (population screened), and higher/lower performance (recall rates, cancer detection rates) to map consistency across multiple NHSBSP units.

Objective 4: Devise a sampling frame based on the survey responses and KC62 breast unit performance data.

Objective 5: Triangulate findings from the literature review, national survey, and interviews to develop the evidence base for how guidance can best be developed and improved for effective use of existing reporting and arbitration processes.

1.5 Thesis Content

This research is informed by a literature review and anecdotal evidence that the majority of breast screening units based in one large region have moved to consensus review in favour of third reader arbitration. Table 2 provides an overview of the thesis and demonstrates how the individual components of the research informed the following elements.

Table 2 Overview of Thesis Content Relative to the Research Objectives

Research Objective	Methods	Chapters
Identify research evidence on the barriers and facilitators for decision-making in mammogram reading. Use a narrative synthesis to develop a conceptual framework. What is the future role of new technologies?	Literature review Exploratory semi-structured interviews	2,3,4 8
Explore current arbitration systems. Variance in practice, the consequences of such variation for imaging professionals, service providers and service user outcomes and the implications of the PHE guidance	Descriptive exploratory online surveys	6
Collect and analyse data on a series of unit characteristics - higher/lower performance (overall recall rates, incident small cancer detection rates, and size of the unit (population screened) to map consistency across multiple NHSBSP units.	Analysis of published KC62 data. Utilise specified performance criteria to develop a sampling frame.	7
Explore current arbitration systems. Variance in practice, the consequences of such variation for imaging professionals, service providers and service user outcomes and the implications of the PHE guidance	Semi-structured telephone interviews.	8
Develop the evidence base on how evidence-based guidelines on reporting/arbitration can best be developed and improved.	Triangulation of findings from the literature review, national survey, unit performance and interview data	9

Chapter 2 of the thesis sets the context of the study providing an overview of the disease of breast cancer, the evidence base relating to breast cancer risks, screening

and diagnosis. The complexity of a breast screening programme is highlighted through discussion of inter-related performance measures, inter-observer variability, emergent technologies, and how arbitration can play a fundamental role in achieving performance standards. The factors influencing breast screening outcomes are critically discussed, emphasising the cognitive complexity of diagnostic decision-making and errors. This provides the rationale for exploring what can be learnt from case studies of decision-making in higher and lower performance units to inform the future effective use of arbitration processes in breast screening. The chapter concludes with the fact that Quality Assurance outcomes of arbitration are not intently reported in contrast to first reader performance, but potentially has the same variability which necessitates consideration.

Chapter 3 is a review of the published literature on clinical decision-making, beginning with a definition, moving on to factors affecting the process, and the theoretical models that underlie decision-making in clinical practice. The chapter concludes with an overview of strategies for reducing error in clinical reasoning and a review of the evidence from medical clinical reasoning studies.

As groups of staff may assess discordant cases, Chapter 4 reviews the published literature on group decision-making and the team-based nature of consensus processes, critically discussing how organisational behaviour and team dynamics can influence the outcome and concluding with how algorithms and Artificial Intelligence (AI) may support the decision-making process.

Chapter 5 discusses the rationale underpinning the design and methodology adopted in this research. The advantages and disadvantages of undertaking a Mixed Method Study are discussed. The priority and sequence of methods are explained along with the phase of the study integration. Ethical considerations are presented, critically reviewing the measures to protect participants from harm. Study quality and rigour are introduced, but as three distinct study phases are undertaken, the detailed methods are described in the respective chapters (6,7 and 8).

Throughout the thesis, the study findings are discussed as they are presented. Chapter 6 presents the first stage of the study in which national online surveys were undertaken. The methodological approach of survey construction is evaluated while critically appraising the data collection method. The chapter provides a systematic analysis of the results from the quantitative responses, with qualitative analysis of the free-text comments.

Chapter 7 presents the second stage of the study, providing an overview of the KC62 data set and Breast Screening Information System (BSIS). A rationale for the chosen performance metrics is presented. These performance metrics are analysed with the survey responses on reporting and arbitration strategies.

The third stage of the study is presented in Chapter 8, which discusses the methodological approach of using semi-structured telephone interviews, offering a rationale for selection while critically appraising the data collection method. The theoretical justification of the study sites and participant sampling strategy are

discussed along with the key research findings.

The final chapter (Chapter 9) compares and contrasts the data with triangulation to provide a succinct summary of the thesis findings. The benefits and limitations of the study design are discussed. Recommendations for practice and potential future projects arising from this research are suggested.

Chapter 2. Background

This chapter aims to consider the main issues and complexity of decision-making within a breast screening setting to explore how discordant reports are reconciled and appropriate action taken. To provide an insight into the complex nature of a screening programme, the concepts of performance measures and inter-observer variability are described, together with the international variance in reporting strategies. After first explaining the aetiology and prevalence of breast cancer, the chapter critically reviews the limitations of current screening using mammography and discusses the role of future technologies. The chapter justifies the reason for exploring variations in reporting and arbitration strategies and the influence on breast screening units within England of the Public Health England (PHE) arbitration guidance.

2.1 Epidemiology of Breast Cancer

Breast cancer represents the most common female cancer worldwide (Worldwide Cancer Data | World Cancer Research Fund, 2018). Figures from the worldwide cancer data confirm that there were 2,088,849 new cases of breast cancer diagnosed in 2018 (represents 25% of all female cancers) and 0.6 million deaths from the disease globally. Figure 1 demonstrates the variance in incidence rates and mortality rates worldwide. Although hereditary and genetic factors account for 5% to 10% of breast cancer cases (Bray et al. 2018), non- hereditary factors are considered the leading cause of the international and interethnic variation in incidence (Ziegler et al. 1993). Higher incidence rates in some countries are

attributed to a higher prevalence of established risk factors which are discussed further in section 2.3 and increases in breast cancer screening and awareness.

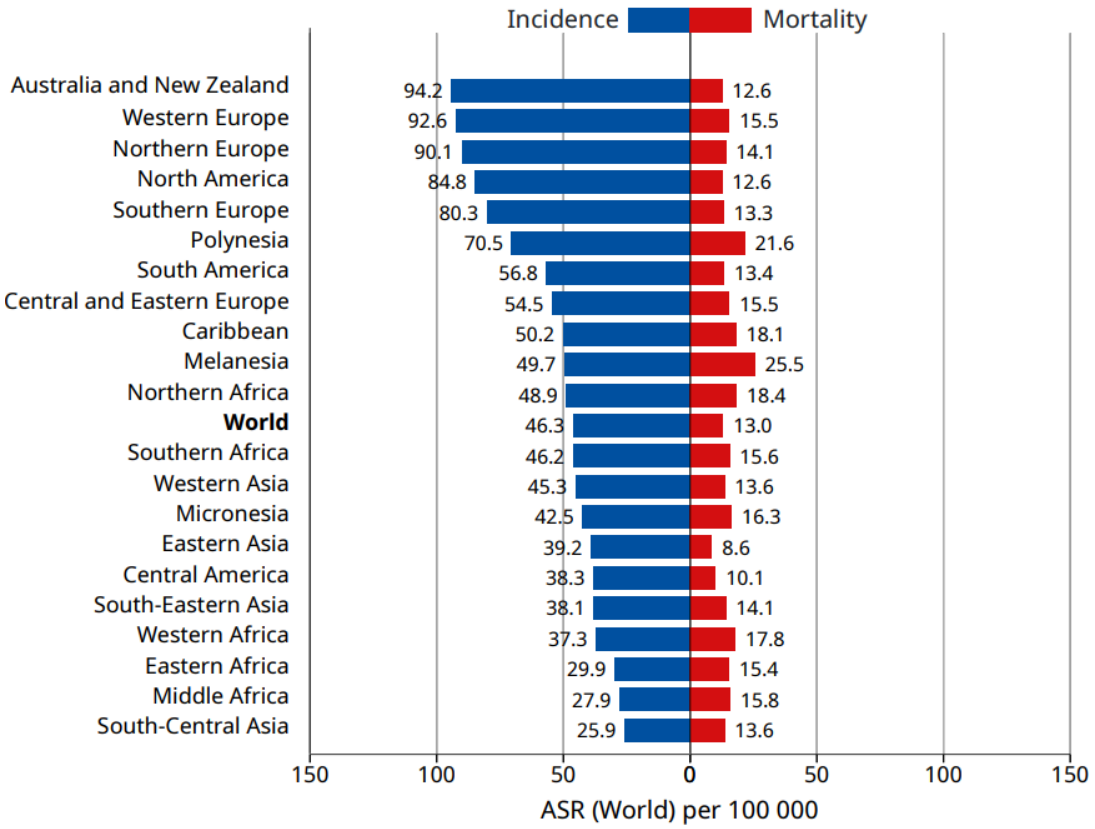


Figure 1 2018 Region-Specific Incidence and Mortality Age-Standardized Rates for Female Breast Cancers.
Source: GLOBOCAN, 2018.

2.1.1 Breast Cancer Prevalence in the UK

In the UK, there were 54,722 invasive female breast cancers diagnosed in 2017 (Cancer Research UK 2020), and there were 11,371 related deaths. The International Agency for Research on Cancer (World Health Organisation 2018) predicts a rise in incidence in the UK from 55 439 cases to 66 612 (11 173 +20.2%) between 2018 and 2040. Reductions in mortality rates are an outcome of improved detection, earlier diagnosis via screening and more effectual treatments delivered

by specialist multi-disciplinary teams (Weedon-Fekjær, Romundstad, and Vatten 2014, Seely and Alhassan 2018).

In recent years, the overall benefit of breast screening has been a subject of debate. The Marmot report evaluated the evidence on the benefits and harms of breast screening from a UK perspective (Marmot et al. 2013). Although the investigation concluded that screening

"prevents around 1,300 breast cancer deaths in the UK per year"

criticism remains regarding excessive false-positive recalls, limited sensitivity, and overdiagnosis (Autier and Boniol 2018). A retrospective comparative study (Újhelyi et al. 2016) of screen-detected and symptomatic cancers reported that screen-detected patients did not show any significant improvement in overall survival (OS) or disease-free survival (DFS) compared to the symptomatic group. However, the tumour size was significantly smaller in the screen-detected group ($P < 0.01$), together with a higher prevalence of negative regional lymph nodes ($P < 0.01$). In the symptomatic cohort, there was a higher incidence of distant metastases (17% compared to 10%) and chemotherapy (17% higher). The study states the data supports a trend of disease-free survival in the screening group but had not reached statistical significance. However, there was only a median follow-up of 65 and 80 months and therefore, a more extended follow-up period possibly will show a statistical significance of DFS and/or OS of the screening patients. A further important consideration mentioned in the study is that symptomatic patients may

require more aggressive treatments, and this may significantly reduce the patient's quality of life.

Currently, population-based breast cancer screening is focussed merely on the age of the woman (50-70, on a three-yearly basis). The exception to this is increased screening for a small cohort of women with a moderate/high increased lifetime risk (>30 %) or high-susceptibility genes (*BRCA1* and *BRCA2).* The AgeX is a nationwide randomised controlled trial to establish if extending the age range further (47–73 years) is beneficial (Moser et al. 2011). The trial commenced in 2009, but the information is not expected until the mid-2020s. The trial received ethical approval for three yearly invitations for ages 71–76 or 71–79 to evaluate the effects of continuous screening after the age of 70. However, routine screening was suspended in the UK in March 2020 due to COVID. The trial investigators, therefore, decided that there would be no further randomisation into the trial as the pandemic has created a considerable backlog on breast screening services.

Screening programmes in other countries generally have a shorter interval between mammograms typically 1- or 2-yearly and additionally may offer to screen from an earlier age, as demonstrated in Table 3. The optimal age range and screening interval remain a topic of debate (Blanks 2011).

Table 3 The International Institutional Variance in Recommended Screening Age and Interval. Taken from the Institutions Listed in the Table (2018)

Institution		**Screening Interval**						***High-risk***	
Society	**Country**	**40-49**	**Interval**	**50-70**	**Interval**	**>70**	**Interval**	**≥40**	**Interval**
American Cancer Society	USA	Offer 40-44 45-54	1 1	Y	1-2	Y	1-2	Y	1
American College of Radiology	USA	Y	1	Y	1	Y	1	Y	1
National Cancer Institute	USA	Y	1-2	Y	1-2	Y	1-2	Y	1
United States Preventative Task Force	USA	Offer or provide the service	1-2	Y	2	Y-74	2	Y	1
National Breast Cancer Screening Programme	Netherlands	N	-	Y	2	Y-75	2	Y	1
Canadian Task Force on Preventative health care	Canada	N	-	Y 50-69	2-3	Y 70-74	2-3	Y	1
Agence Nationale d'accreditation et d'Evaluation en Sante	France	N	-	Y	2	Y-74	2	Y	1
National Health Breast Screening Programme	UK	N routinely AgeX trial (47-73)	-	Y	3	Y	3	Y	1
Swedish National Board of Health and Welfare	Sweden	Y	18 months	Y	18-24 months	Y	18-24 months	Y	1

2.2 Breast Cancer Aetiology and Pathology

Numerous genes are involved in controlling the process of normal cell division. This process requires an equilibrium of activity between the genes that stimulate and suppress cell proliferation and those that signify when damaged cells should undergo apoptosis, which is a form of controlled cell death. Once mutations

accumulate in the genes responsible for cell proliferation, cancerous cells develop (Broustas and Lieberman 2014).

Breast cancer is a heterogeneous disease with multiple subtypes (Sinn and Kreipe 2013). The majority of breast cancers develop from epithelial cells and may be *in-situ* disease (pre-invasive) or invasive disease. At a pre-invasive (*in-situ*) stage, the malignant cells have not breached the basement membrane surrounding the ducts and lobules. *In-situ* disease is further classified as ductal carcinoma in situ (DCIS) or lobular carcinoma *in situ* (LCIS), with further subdivision by nuclear grade and architectural features. High-grade DCIS is deemed to represent a higher risk of progression to invasive cancer. Low and low-intermediate grade DClS are now considered lower risk, and this represents the hypothesis of the current breast screening LOw RISk DCIS trial (LORIS) (Francis et al. 2015). The purpose of the trial is to determine if the historical practice of surgical excision (local excision or mastectomy) of low-risk DCIS is substantiated.

LCIS can be a difficult disease to manage, as it is more likely to be multifocal, multicentric and affect both breasts (bilateral disease). LCIS termed 'classic', represents a very low risk of progression to invasive cancer over 25 years (Stewart et al. 2014). However, LCIS classified as 'pleomorphic' is considered a higher-grade variant and the rate of progression of this is uncertain at present. DCIS detected via screening is usually asymptomatic, depicted as minute deposits of micro-calcification on a mammogram. The introduction of breast screening led to a substantial rise in the incidence of DCIS/LCIS and hence, the concerns of over-diagnosis have been discussed extensively in the literature and lay press.

Consequently, women undergo treatment for a low-grade disease that may be considered unnecessary, as if left undiagnosed and untreated; it may never progress to invasive cancer (Gøtzsche 2012).

Morphological features also characterise invasive breast cancers. The majority (80%) of invasive cancers are from the heterogeneous group of ductal carcinomas (no specific type)(Sandhu et al. 2010). The most frequent of the special subtype is lobular carcinoma (10%). The less common subtypes include mucinous, tubular, medullary, cribriform, micropapillary, papillary, metaplastic, and inflammatory carcinomas (Sandhu et al. 2010).

Three tumour parameters are currently used to grade invasive cancers:

I. Tubule formation
II. Nuclear pleomorphism
III. Mitotic count

The histological grade determines how similar a tumour is to the tissue of origin. Table 4 demonstrates the scoring calculation, and the final grading used as an indicator of patient prognosis.

Table 4. The World Cancer Report (2014)

Semi-quantitative method for assessing histological grade

Feature	Score
Tubule and gland formation	
Majority of tumour (> 75%)	1
Moderate degree (10–75%)	2
Little or none (< 10%)	3
Nuclear pleomorphism	
Small, regular, uniform cells	1
Moderate increase in size and variability	2
Marked variation	3
Mitotic count	
Dependent on microscope field area	1–3
Final grading	
Add scores for tubule and gland formation, nuclear pleomorphism and mitotic count:	
Grade 1	Total score 3–5
Grade 2	Total score 6 or 7
Grade 3	Total score 8 or 9

The histopathological analysis also enables assessment of disease stage, which is determined by tumour size and regional lymph node involvement. The overall staging for the patient is recorded using the tumour–node–metastasis (TNM) classification (Appendix 2). Supplementary prognostic information regarding lymphovascular invasion and levels of response to neoadjuvant treatment can also be collated from the histopathological assessment. However, limitations are acknowledged in this pathological assessment for managing breast cancer. Over the last decade, substantial advances have been made in comprehending the biology of breast cancer and translating some of the molecular data to inform clinical care. DNA microarray technology for gene expression profiling has been employed to classify breast cancers, develop indicators of good and poor prognosis tumours, and

classify those that may or may not respond to specific therapies (van de Vijver et al. 2002, Sandhu et al. 2010).

2.3 Risk Factors

2.3.1 Age

The most influential risk factor for breast cancer in women is age, with increasing age the higher the risk. Figure 2 from Cancer Research UK (2015-2017) demonstrates a plateau after 50-54 when breast screening is first routinely offered and relates to the detection of prevalent cases.

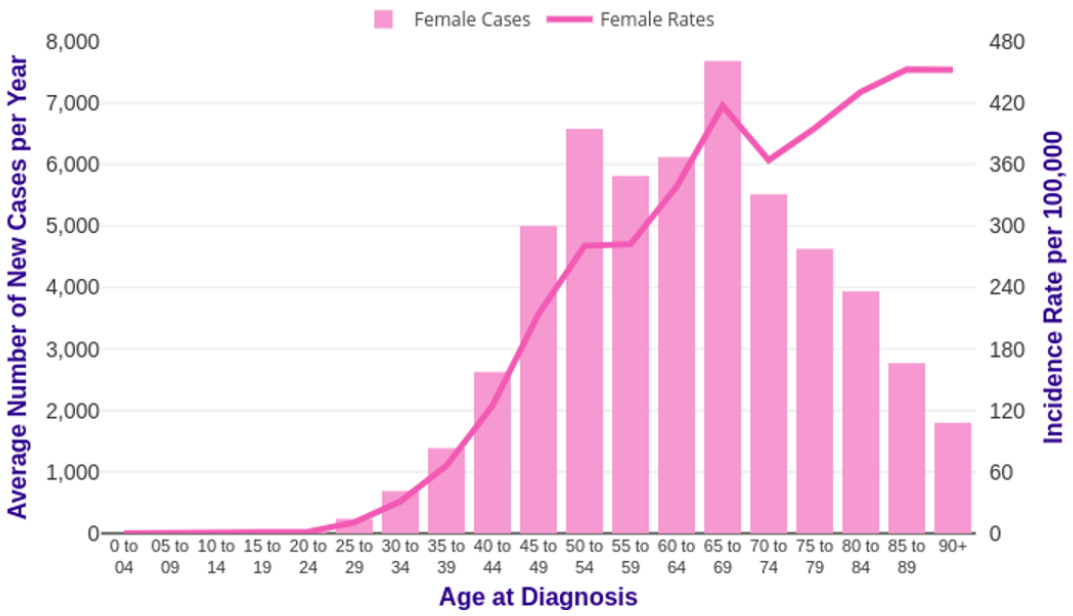


Figure 2 Demonstrating the Average Number of New Female Breast Cancer Cases per Year Relative to Age, UK, 2015-2017
(Taken from Cancer Research UK)

2.3.2 Endocrine/Reproductive/Lifestyle Factors

Multiple risk factors are associated with increasing an individual's probability of developing the disease (Stewart et al. 2014). These involve endocrine and reproductive factors; nulliparity (never borne a child), age at first birth, age at

menarche and menopause; exogenous hormone intake (hormone replacement therapy or oral contraceptives use); lifestyle factors relating to nutrition, alcohol consumption, anthropometry, physical inactivity and exposure to ionising radiation. Barnes et al. (2011) state that for invasive cancers the population risk for non-variable factors (menarche and menopause age, family history of breast cancer, and personal history of benign breast disease) was 37.2% (27.1–47.2%, 95% CI). HRT (19.4%; 15.9–23.2%, 95% CI) and lack of exercise (12.8%; 5.5–20.8%, 95% CI) were attributed to the highest population risks of modifiable factors. These results varied depending on the hormone receptor status of invasive cancers.

2.3.3 High-Risk Groups

The majority of breast cancers are not familial. A small percentage of women are identified as having a high risk, a consequence of mutations in the BRCA1 and BRCA2 (breast cancer susceptibility) genes (Ford et al. 1998). Mutations in these genes are classified as high-penetrance with an average relative risk of 11.4 and 11.7 respectively, along with the rare gene TP53 (Li-Fraumeni syndrome) with an age-adjusted relative risk 105 (90% CI 62–165) (Wendt and Margolin 2019). Moderate risk genes, for example, CHEK2, ATM, PALB2 and RECQL confer average relative risks of 2.26-5.3 (Wendt and Margolin 2019), with further low-penetrance gene variants also identified (Turnbull and Rahman 2008).

2.3.4 Breast Density

The mammographic density (MD) of a breast relates to the breast parenchyma that is denser than the adipose (fatty) tissue. MD can be assessed qualitatively, being a visual observation of parenchymal patterns/distribution using one of the classification systems (Wolfe 1976, Gram, Funkhouser, and Tabár 1997, D'Orsi 2013).

Table 5 demonstrates the latest American College of Radiology (ACR) breast density classification.

Table 5 The American College of Radiology BI-RADS Atlas Fifth Edition – Breast density classification (Taken from D'Orsi et al. 2013).

Breast Composition	a. The breasts are almost entirely fatty
	b. There are scattered areas of fibroglandular density
	c. The breasts are heterogeneously dense, which may obscure small masses
	d. The breasts are extremely dense, which lowers the sensitivity of mammography

However, these classifications are subjective with varying reproducibility of results (Vinnicombe 2018). Quantitative methods may also be visual as in the BI-RADS 4th edition which categorised density into percentages (0-24%; 25-49%; 50-74%, and 75%) (D'Orsi 2013) or the Boyd six category classification (Boyd et al. 1995) and visual analogue scales (Duffy et al. 2008). There are also validated semi-automated methods. Several studies have shown these systems to correlate well with breast cancer risk (Boyd et al. 2007, Eng et al. 2014). However, as they require user input, they are considered impractical clinically (Vinnicombe 2018). Automated volumetric methods have demonstrated consistency in density grading and have also shown a good correlation with breast cancer risk (Eng et al. 2014, Brand et al. 2014).

Various studies have established that women with a high-density breast tissue (> 75% glandular tissue) have a higher risk (4-6 times) of developing a breast carcinoma, comparative to those with a low breast density (<5 % glandular tissue) (Winkel et al. 2016 and Zhang et al. 2018). A recent study (Engmann et al. 2017) has affirmed that

high breast density was considered the predominant risk factor equally for premenopausal and postmenopausal women representing the main effect on population-attributable risk proportion of breast cancer. The other main problem associated with a high breast density is that the mammographic sensitivity is reduced, as dense glandular tissue conceals the detection of tumours and therefore the risk of an interval carcinoma is greater (Boyd et al. 2007). Destounis et al. (2017) report a linear association relating to mammographic sensitivity and breast density, with sensitivity decreasing from 95% in a fatty breast to 65% in an extremely dense breast. This predicament has been a source of much of debate (Onega et al. 2014 and Schousboe et al. 2011), with personalised screening based on risk factors and mammographic density assessment considered to be the future. Legislation laws have been put into effect in 38 states in the USA so that women who have undergone mammography are informed of their breast density and associated risk (Vinnicombe 2018). Computer software analysis, although capable of measuring breast density, is still being evaluated to assess the consistency of results relative to density changes over time (Oliver et al. 2015). Therefore, the Australian Standing Committee on Breast Screening states that until more evidence is available on breast density assessment, management and clinical pathways, routine recording of breast density and supplementary screening would not be undertaken. Currently, there is no requirement within the NHSBSP to record breast density. Public Health England is evaluating this issue as there are increasing pressures from clinicians and patients (Sharma 2018).

2.4 Breast Cancer Screening

A population-based screening programme can only be effectual if specific measures are sustained, such as adequate coverage and uptake, high sensitivity and specificity with a resultant low rate of interval cancers and false-positive screens. Table 6 demonstrates the possible results following a screening mammogram.

Table 6 Possible Results of a Screening Test

	Cancer Outcome	
	+ Patient confirmed to have breast cancer	- Patient does NOT have breast cancer.
Test + Cancer suspected. based on the mammogram	True positive (TP)	False-positive (FP)
Result - Normal mammogram	False-negative (FN)	True negative (TN)

Many factors contribute to the quality of a mammography programme, including the knowledge/skills and experience of the staff, the imaging equipment used, and the organisation of service delivery at a given breast screening unit (Hopkins 2011).

2.4.1 Detection and Diagnosis

Mammography is currently the technique for population-based screening. Full-field digital mammography (FFDM) was a significant evolution over screen-film mammography (SFM), providing consistent higher contrast resolution images with a lower radiation dose (Juel et al. 2010). The results of the Digital Mammographic Imaging Screening Trial (Pisano et al. 2008) established that FFDM revealed a

substantial improvement for imaging younger women (< 50 years) and those with mammographically dense breasts. However, conventional 2-Dimensional (2D) digital mammography still has limitations, mainly the overlapping of glandular breast tissue which may obscure underlying lesions or mimic a significant finding (Laming and Warren 2000).

2.4.2 Mammographic Interpretation of Images

Mammographic images are examined for abnormalities in the form of masses, microcalcifications, asymmetric densities, and architectural distortions. However, due to a plethora of normal variants and some overlap of features associated with benign and malignant lesions, reporting mammograms remains challenging (Heywang-Köbrunner, Hacker, and Sedlacek 2011). Approximately 10–15% of cancers in women of screening age are not visible on mammography. The UK based TOMosynthesis with digital MammographY (TOMMY) trial (Gilbert et al. 2015) stated FFDM has a sensitivity of 87% but a specificity of only 58%. The specificity of screening mammography can be lower on initial screening examinations but may increase to 93% or higher on subsequent screens, when previous images are available for comparison. Visibility of a lesion on mammography is dependent upon several confounding factors; image quality, the tumour type, the breast density as previously discussed and inter-observer variability. Certain tumour types exhibit minimal mammographic changes and can be extremely subtle to visualise, particularly lobular carcinomas. Even large tumours may be occult (not visible on mammography) if there is minimal effect on the surrounding breast parenchyma.

2.5 NHSBSP Standards/Performance Measures

The NHSBSP developed quality standards to ensure that local programmes are safe and efficient. The performance of a breast screening unit is measured by meeting specific indicators that relate directly to patient outcomes (Appendix 3). KC62 is a statutory annual return completed by individual screening units that record activity and outcome data in the NHSBSP. Table 7 demonstrates the definitions of what the NHSBSP classify as an acceptable and achievable threshold for services to attain.

Table 7 The NHSBSP Classifications of Acceptable and Achievable Thresholds (taken from NHS Breast Screening Programme Consolidated Standards Public Health England 2017)

Acceptable threshold
Is the lowest level of performance services are expected to attain to ensure patient safety and service effectiveness. All units are expected to exceed the acceptable threshold and to agree on service improvement plans that develop performance towards an achievable level. Programmes not meeting the acceptable threshold are expected to implement recovery plans to ensure rapid and sustained improvement.
Achievable threshold
Represents the level at which the services are likely to be running optimally; screening services should aspire towards attaining and maintaining performance at this level.

The difficulty is achieving the appropriate balance between high detection (sensitivity) of early-stage disease while limiting false-positive findings that cause unnecessary further tests, patient anxiety, and additional cost (time and resources of staff) (Welch and Passow 2014). NHSBSP standards (PHE 2017) pertinent to this research are Standards 9, 11, 15, and 16 (Table 8) as they are associated with maximising the number of cancers identified (standard 15) and detecting the cancers at an early stage (standard 16). Standard 9 is of particular importance as

arbitration of discrepant reads is influential in reducing recalls of false positive cases. A systematic scoping review undertaken by (Hackney et al. 2017) highlighted that some units were reviewing concordant recalls in an attempt to lower benign recalls. However, units are required to reach a definitive diagnosis promptly to ensure that results are received within the NHSBSP standard (8) of two weeks from attendance for the mammogram or recalled to an assessment clinic within three weeks (standard 11). Early stage cancers can present as small, subtle lesions with a minimal mammographic change from prior imaging and hence the decision-making on discrepant cases are particularly demanding. The process of arbitration is, therefore, paramount to ensure cases are rigorously evaluated and minimise the risk of a cancer case presenting between screening episodes (standard 19).

Table 8 Selected National Performance Thresholds for the NHSBSP (Taken from PHE 2017).

Objective	Criteria	Performance thresholds	
		Acceptable	**Achievable**
9. To minimise the number of women screened who are referred for further tests whilst trying to minimise false negative rates.	The proportion of eligible women with a technically adequate screen who are referred for assessment	Prevalent screen <10% Incident screen <7%	Prevalent screen <7% Incident screen <5%
11. To minimise the interval from the screening mammogram to assessment	The percentage of women who are offered an appointment at an assessment centre within three weeks of attendance for the screening mammogram	>98%	100%
15. To maximise the number of cancers detected	The SDR is the ratio of the observed number of invasive cancers to the expected number in the eligible population invited and screened	1.00	1.40
16. To maximise the number of small invasive cancers detected	The standardised detection ration (SDR) is the ratio of the observed number of invasive cancers to the expected number in the eligible population invited and screened. Small cancers (<15mm in diameter) should be 55% of the expected overall number of invasive cancers.	1.00	1.40
19. To minimise the number of interval cancers presenting between screening episodes.	The number of interval cancers per 1000 women screened	<0.65/1000 diagnosed <12 months of the previous screen <1.40/1000 diagnosed between 12 and <24 months of the previous screen <1.65/1000 diagnosed between 24 and <36 months of the previous screen	Achievable: n/a Analysis of interval cancer data should take place at screening service level aggregating several years performance

2.5.1 Coverage/Uptake Rates

Coverage is classified as

"*The percentage of women in the population who at a particular point in time are eligible for screening and have a recorded result within the last three years"* (Public Health England 2017:12).

Coverage incorporates women routinely invited, self-referred or referred via their GP. The 2020 NHS Digital data (data 2018-2019) demonstrates that national coverage (women 53-70) fell to 74.6% from 74.9% in the previous year but remains above the NHSBSP acceptable level of 70%. However, the uptake rates (women who attend for screening within six months of invitation) vary by regions. The North East reported the highest uptake (75.3%). London reported the lowest uptake (64.0%) and the North West (69.5%) was also below the acceptable level of 70%. The uptake rates are fundamental if breast screening is to remain effective in reducing mortality from breast cancer. However, it may be challenging to achieve dependent on local population demographics. Deprivation and high populations of certain ethnic minority groups are associated with lower uptake (Massat et al. 2015).

2.5.2 Recall Rate

The recall rate is defined as

"*The number of screened women recalled for further assessment as a proportion of all women who had a screening examination"* (PHE 2017: 18).

Recall rates to assessment are reported by prevalent (first invitation for screening and routine invitations to previous non-attendees) and incident (routine invitations to previous attendees screened within five years) status. It is envisaged that recall rates are lower for incident screens as, only new disease that has developed since the last screening mammogram will be detected. The NHSBSP define the acceptable (<10% Prevalent screen and <7% Incident screen) and achievable thresholds (<7% prevalent screen and <5% incident screen) for recall rates (standard 9 Table 8). However, as identified in a systematic scoping review (Hackney et al. 2017), there is international variance in the achievable standards with lower European guidelines (<5% for prevalent screens and <3% for incident screens). The Dutch Screening Programme reports the lowest recall rates worldwide averaging 1.6% with the American College of Radiology recommending an overall recall rate of <10% (USA <12%). However, recall rates are not comparable internationally, due to the variance in the recommended screening age range and time interval as discussed previously.

It is established that a correlation exists between recall rates and early detection of breast cancers (Otten et al. 2005). However, what may be an 'optimal' recall rate remains a source of debate. High recall rates would infer that a unit is over-recalling women who will undergo additional assessment (false positive) and is an inefficient use of a service's resources. Conversely, low recall rates may result in lower cancer detection rates. The complexity of this concept is summarised in Table 9 (D'Orsi and Sickles 2017).

Table 9 Demonstrating the Potential Relationship Between Recall Rates, Cancer Detection Rates and Positive Predictive Value (PPV)
(Taken from D' Orsi and Sickles 2017).

(a) A screening recall rate at or below the benchmark level associated with a low cancer detection rate (CDR) indicates poor performance because the primary goal of screening is early detection.
(b) A PPV substantially higher than the benchmark level may indicate poor performance because, in this scenario, only those lesions with a much greater probability of malignancy are considered actionable, forgoing earlier detection of the subtler albeit less specific malignancies.
(c) A below benchmark screening recall rate coupled with above-benchmark CDR indicates optimal performance.

Yankaskas (2004) states that variations in recall rates are not fully comprehended and maintains this may be a result of differences in programme constitution, variance in recall definition and data collection. This may be justifiable for international differences but would not explain the variation in recall rates across England in established programmes operating under the same NHSBSP guidance.

Notably, historical literature discussed recall rates as a single measure. Otten et al. (2005) assert that the correlation between recall rate and cancer detection rates are complicated; a view supported by Mohd Norsuddin et al. (2015) who depicted this in a conceptual framework. This framework has been modified (Figure 3) to represent UK practice and standards and to also highlight that error occurs in interpretation, not just in perception. Emergent technologies are included, and the framework also demonstrates where arbitration sits within the multifactorial nature of recall rates.

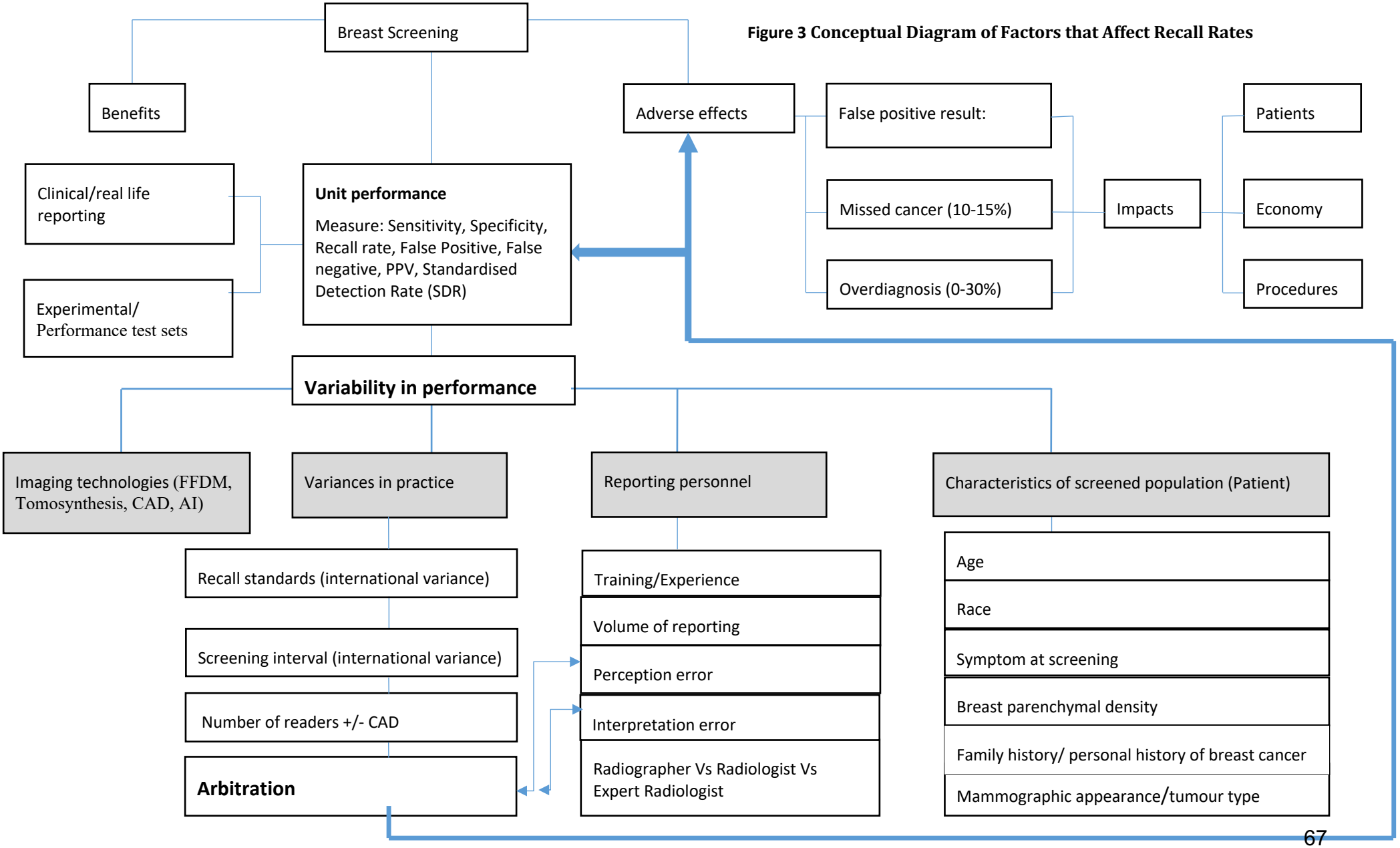


Figure 3 Conceptual Diagram of Factors that Affect Recall Rates

Table 10 (Appendix 4 Studies exploring the association between recall rates and performance measures) summarises the empirically based studies that have considered the association between programmes recall rates and performance measures. The study undertaken by Gur et al. (2004) claimed a statistically significant linear fit between recall and cancer detection rates. However, only a small number of experienced Radiologists were involved in the study. Nevertheless, it did demonstrate a substantial variation between individual readers for both measures (recall rates ranged from 7.7% - 17.2% and CDR 2.6 - 5.4 per 1000 mammograms). Conversely, some researchers advocate that there is not a strong correlation between recall rates and cancer detection, particularly above a certain threshold (Yankaskas et al. 2001a and Otten et al. 2005). Recall rates in the USA are reported to be higher, and malpractice concerns may be a contributory factor in particular screening programmes (Otten et al. (2005). A recent USA study undertaken by Grabler et al. (2017) maintains that a recall range of 12%- 14% would provide optimal cancer detection rates. Previously radiology groups in the USA have considered double reporting as labour-intensive and not cost-effective. However, Mullen et al. (2017) advocate that time efficiencies saved from recalls and subsequent workup could be offset against the second reading of all screening cases to improve quality.

The NHSBSP guidelines (Public Health England 2017a) acknowledge that screening units may not always strive to reduce recall rates dependent upon cancer detection rates. In services with especially high cancer detection rates reducing referral to assessment rates may not be realistic. The NHSBSP also states that recall rates will

vary with experience of the readers; experienced readers are likely to have lower recall rates on average than inexperienced readers. The emphasis on recall rates in this current study pertains to the belief that arbitration can play a fundamental part in units recall rates. Ideally, arbitration can decrease the number of false-positive findings while maintaining (or improving) cancer detection rates, therefore sustaining the overall purpose of screening. However, regional data presented at the *Symposium Mammographicum Conference 2016* (Steel) confirm that third person arbitration results vary widely depending upon the individual undertaking the task. Significant differences were reported in the proportions of cases going to arbitration, cases recalled, and cancers detected following arbitration (all $p < 0.001$).

2.5.3 Cancer Detection Rate (CDR)/Age Standardised Detection Ratios (SDR) for Invasive Cancers

Evaluating cancer detection rates between breast screening units is imperative. The revised 2017 PHE Consolidated Standards for NHSBSP has withdrawn reporting of invasive cancer detection rates (CDR) and replaced it with standardised detection ratios (SDR). CDR is no longer considered valid in England due to the variability in the mean age of women screened. This is a result of some units participating in the age extension trial. As discussed previously, age is a significant risk factor affecting cancer detection rates, and the SDR allows for correction of the age distribution of the eligible population invited and screened by comparing the observed invasive cancers to the expected number of invasive cancers. The ratio is based on measures from the Swedish Two Counties randomised control trial (Tabár et al. 1985), which is

utilised as the performance comparator. An SDR of 1.00 represents equivalence with this trial. NHSBSP Standard 15 in Table 8 states that 1.00 is acceptable with 1.40 an achievable performance threshold. Small cancers are those classified <15mm in size and should represent 55% of the expected overall number of invasive cancers (Standard 16 NHSBSP). However, caution must be applied when comparing performance rates for small units relative to large units. It is acknowledged that the frequency of screen-detected cancers is low, and therefore the yearly CDR/SDR for small services may lack sufficient statistical accuracy to be meaningful.

2.5.4 Interval and Missed Cancers

Cancers that develop between scheduled screening episodes (3 years in the NHSBSP) are termed interval cancers. Evans et al. (2016) state that 40% of tumours develop in this 3-year interval. These cancers are associated with a worse prognosis and subsequently can reduce the potential effectiveness of breast screening (Howell et al. 2005). Interval cancers can include fast-growing tumours becoming mammographically detectable and clinically apparent after the screen. Also, they may exist, but are mammographically occult (not visible- a limitation of mammography) or have been missed by the reporter(s) (Heywang-Köbrunner, Hacker, and Sedlacek 2011). Interval cancers are classified into three categories, as demonstrated in Table 11.

Table 11 NHSBSP Classification for Interval Cancers (Taken from Public Health England 2017b).

	Category	Radiological	Action Warranted	Disclosure of Audit/Duty of Candour
1	Satisfactory	Normal, benign mammographic features	No reason to recall	Disclosure of audit
2	Satisfactory, with learning points	Difficult to perceive, seen with hindsight, not obviously malignant	Not all readers would recall, may provide some learning	Disclosure of audit
3	Unsatisfactory	Appearance is obviously malignancy	Should have been recalled, all readers reviewing the films agree that they would recall	Classify as a notifiable safety incident under Duty of Candour process

The NHSBSP aims to minimise the number of interval cancers. The 2017 PHE guidance (Public Health England 2017a) deemed a revision of interval cancer thresholds (Standard 9) was required to correspond with the natural increase in incidence (25% from 1995) of breast cancers. Previously performance thresholds were reported at <24 months and 24-36 months. They are now divided into three values to reflect each subsequent year following a normal screening. It is acknowledged that yearly interval cancer rates are small in individual units and therefore PHE state that analysis should be undertaken on an accumulation of several years' performance. It is also essential that they are not analysed in isolation from other performance data, specifically SDR.

2.6 Factors Influencing Breast Screening Outcomes

2.6.1 Double Reporting

Double reading has been implemented in many European countries (e.g. Sweden, Hungary, Netherlands) to increase cancer detection rates and minimise reporter

error. However, a systematic scoping review (Hackney et al. 2017) highlighted that there is international inconsistency in how screen reporting is conducted. There is variance in the professional roles undertaking breast screen reporting. The European standard is double reporting by Radiologists specialised in breast screening. In the United States, single Radiologist reporting or single Radiologist reporting with Computer-Aided Detection (CAD) is the norm. Breast Clinicians are also utilised in Australia and the UK, but exclusive to the UK is double reporting by Radiographers. When the NHSBSP was founded, only medically qualified professions (Radiologists, Clinicians) were eligible to interpret and report the mammographic images. A subsequent shortage of breast Radiologists necessitated a change in service delivery. In the UK, Radiographers were formally trained to report screening mammograms, and Pauli et al. (1996) confirmed that Radiographer reporting was as accurate as that of Radiologists. A further progression occurred in 2012 following an extensive research project (NDROR) (Bennett et al. 2012) which endorsed double Radiographer reporting. The success of role extension in the UK has preceded international researchers to consider training mammographers in the reporting of mammograms (Debono et al. 2015, Torres-Mejía et al. 2015, and Moran and Warren-Forward 2016).

2.6.2 Blinded vs Non-Blinded Reporting

The systematic scoping review (Hackney et al. 2017) also identified variance with regards to reading practices. Some services utilise true blind reading (the second reader is not aware of the first reader's decision on the computer software or assessment paperwork); in other units, the second reader is blinded to the first

reader's decision on the computer software but can see the final report by looking at the assessment paperwork. Alternatively, non-blinded (the first reader's decision is available on the computer screen) reading occurs. Klompenhouwer et al. (2015) identified that there is a dearth of studies comparing the advantages and disadvantages of blinded vs non-blinded reporting and concluded that blind reading increases a unit's sensitivity but generates more discrepant cases. It is currently unknown what strategies units within England are utilising, and the value of second reading could be questioned if not blinded.

2.6.3 Resolving Discordant Readings

Discordant readings may be resolved either by the two reporters discussing and attempting to reach an agreement; or referring to a single third reader or group of reporters for evaluation. Consensus approaches encompass a diverse range of scenarios, and it was not possible from the scoping review to establish the rationale for the variance or how consensus meetings could be optimally structured. In particular, no studies were retrieved, which evaluated the influence of the component factors, e.g. group structure, group size, group dynamics within a hierarchical structure in a clinical environment. Complex pathways were also described in the literature where both group consensus and third reader arbitration are undertaken, or decision processes weighted by the initial scoring (level of suspicion) of the lesion.

Each system requires differing amounts of personnel and time, with resultant differences in cost. Double reading with a consensus of concordant and /or discordant cases would be one of the costliest and personnel-intensive approaches.

It has been suggested that interval cancer rates can be considerably higher in cases that have undergone arbitration or consensus review comparative to rates of concordant normal screens (Jenkins et al. 2014 and Hofvind et al. 2009). This raises the question of whether arbitration could be refined to aid earlier detection in such cases.

2.6.4 Reader Performance

Numerous studies have reported that the interpretive acumen of mammography reporters is hugely variable (Miglioretti et al. 2007, Elmore et al. 2009, Skaane et al. 2008, Duijm et al. 2009, Lehman et al. 2017, Giess et al. 2019). Contributory factors include the low frequency of screen-detected cancers; decision-making in complex clinical settings and the uncertainties associated with human decision-making. The elements of the NHSBSP guidelines (Hopkins 2011) which relate to interpretive performance include: (1) formal audit; (2) requirements related to initial training, maintaining knowledge and Continuing Medical Education (CME)/ continuing professional development (CPD); (3) interpretive volume (minimum of 5000 screening and/or symptomatic cases per year); (4) participate in PERFORMS (Personal Performance in Mammographic Screening) and (5) participate in screening assessment and MDTs.

The volume of mammograms reporters are required to interpret per year varies internationally ('Mammography Quality Standards Act' 1992, Perry et al. 2007, Public Health England and PHE 2011, European Commission Initiative on Breast Cancer 2019). Some researchers maintain that there is inconsistent evidence between the numbers reported and recall/cancer detection rates, but this may be

attributable to the many other variables amongst international programs (Buist et al. 2011a, Théberge et al. 2014, Duncan and Scott 2011). However, several studies have shown a stronger correlation between increased volume and lower false-positive rates, (Buist et al. 2011, Perry et al. 2007 and Hofvind et al. 2008) and those increasing years of experience correlated with decreasing recall rates (the inverse is true for PPV) (Miglioretti et al. 2009). A regional Film Reader Quality Assurance (FRQA) performance report, which is based on data as a first reader (data 2012-2015) demonstrates that the group of reporters who were reading <10,000 mammograms over the three years had significantly higher recall rates, lowest PPV and the highest discrepant cancer rate. Those who read 20,000-25,000 cases had the highest cancer detection rate. Reporters reading >25,000 had a significantly lower recall rate, the highest PPV and no decrease in discrepant cancer rate. The report concluded that the differences were not significant but demonstrated a trend towards an improved performance for those reading more than 10,000 cases over three years, and therefore substantiating the minimum number of reads advised by the NHSBSP.

Onega et al. (2014) found that centres with higher volumes of reporting were notably detecting small (<15mm) early-stage (and lymph node-negative) disease compared to those with lower volumes. Many of the NHSBSP guidelines pertinent to reporting have been transposed into the recent PHE arbitration guidance (PHE 2016).

2.7 Human Error

2.7.1 The Nature of Diagnostic Errors

Diagnostic errors can adversely impact on patient well-being leading to adverse health outcomes, psychological distress, and financial costs (Singh et al. 2017). The World Health Organization (WHO) acknowledged the significance of errors in diagnosis (Cresswell et al. 2013). Diagnostic images represent raw data, which the reporter has to interpret via processes of visual detection, pattern recognition, memory exemplars and cognitive reasoning. These processes are influenced by the individual's knowledge, experience and cognitive biases (Brady 2017), and diagnostic reports are, therefore, a subjective interpretation.

2.7.2 Errors in the Context of Screen Reading

Errors in reporting can occur as there is a requirement for prolonged periods of concentration, with a requirement to report quickly. A recent systematic review (Stec et al. 2018) concluded that these factors contribute to fatigue with eyestrain and blurred vision intensifying relative to the number of images reported. Breast screen reporting is a repetitive task, and in UK practices where large volumes of films may be read sustaining focus may be difficult, leading to fatigue and affecting diagnostic accuracy. Interestingly, a large randomised clinical trial in breast screen reporting was conflicting with a vigilance decrement not being observed (Taylor-Phillips et al. 2016).

A false-negative (missed cancer) report is either due to a perceptive error or interpretative error (Cornford et al. 2005). Yankaskas et al. (2001b) and Hoff et al. (2011) report that almost one-third of cancers are visible retrospectively on the

previous mammography images. Even with vigilance and experience, perception errors can still occur due to the non-specific features of certain lesions (i.e. seen on one view only, low-density lesions, developing asymmetries particularly within dense breast tissue) (Goergen et al. 1997). Interpretative errors occur when an abnormality is identified but misinterpreted. This may be a result of a knowledge deficit or cognitive bias. Wadhwa et al. (2016) state that misinterpretation is often associated with microcalcification, well-defined masses and progressive asymmetries. Stability of a lesion can also represent a pitfall as this does not always equate to benignity. Slow-growing, low-grade tumours may not show any or only minimal change over a period of time. Bankier et al. (2010) state that even when two readers agree or achieve consensus on an image, this does not necessarily equate to a correct decision.

In the NHSBSP regular audit and review of personal and team results are mandatory. However, the outcomes of third reader arbitration and group consensus are not as closely studied or reported as first and second reader performance. The inconsistency in third reader performance has not been depicted and may have a considerable effect clinically (Steel 2016). Third reader variability requires this same level of attention.

"Quantitative guidelines may be helpful for new arbitrators in the NHSBSP" (Steel 2016*).*

A culture of reflective learning by reviewing interval cancers and screen-detected cancers aims to provide feedback on diagnostic performance in an attempt to

improve diagnosis and reduce errors. However, Berenson et al. (2014) and Croskerry (2012) also assert that poorly understood characteristics of the diagnostic and clinical reasoning processes, also contribute to error.

2.8 Emergent Technologies that may Improve Cancer Detection

Although the Digital Mammography Screening Trial (DMIST study) (Pisano et al. 2008) demonstrated Full-Field Digital Mammography (FFDM) improved sensitivity in younger women (less than 50 years) and those with mammographically dense breasts, it did not show an overall sensitivity improvement compared to film-screen mammography. Two further technologies, Digital Breast Tomosynthesis (DBT) and Contrast-Enhanced Spectral Mammography (CESM) have since been developed.

2.8.1 Digital Breast Tomosynthesis (DBT)

Digital Breast Tomosynthesis image acquisition results in multiple reconstructed thin slices, thereby minimising the problem of overlying breast structures associated with two-dimensional (2D) mammography. Evidence has been accumulating for DBT as a supplementary screening tool to 2D-mammography or a stand-alone technique. DBT aims to increase the detection of invasive cancers while simultaneously reducing false-positive results (Friedewald et al. 2014). Results from DBT trials and observational studies have demonstrated differing results, some reporting lower false-positive rates with the use of DBT compared to DM, while others have suggested higher rates (Bernardi et al. 2016, Lowry et al. 2020, Skaane et al. 2013a, Lång et al. 2016, Friedewald et al. 2014, Sankatsing et al. 2020).

Multiple studies report reductions in recall rates when comparing DBT with DM. Two prospective single-site European screening studies (Skaane et al. 2013 and Ciatto et al. 2013) reported a 15% and 17% reduction in recall rates. Two observational single-site studies in the USA (Rose et al. 2013 and Haas et al. 2013) demonstrated substantial reductions in recall rates of 37% and 30%, respectively. However, a 2018 meta-analysis undertaken by Marinovich et al. (2018) stated that reductions in recall rates were mainly found in USA studies. The USA has a higher baseline recall rate comparative to European countries and, therefore, the decrease in recalls and false-positive results are dependent on the initial DM rate. A UK prospective randomised study (Maxwell et al. 2017) reported that the addition of DBT to 2D-mammography in incident screening did not show a significant reduction in recall rates, but may increase indecisiveness until reporter experience is developed.

The evolving literature shows that DBT increases detection of invasive breast cancers compared with DM alone (Houssami and Miglioretti 2016, Ciatto et al. 2013, Marinovich et al. 2018, Durand et al. 2015, Rose et al. 2013, Friedewald et al. 2014, Skaane et al. 2013b). However, there is currently a dearth of evidence to establish if DBT reduces breast cancer mortality or the effect on potential overdiagnosis (Welch and Passow 2014, Hovda et al. 2020). Outcome measures such as interval cancers are required, and the results from the prospective population-based Oslo Tomosynthesis Screening Trial (OTST) (Skaane et al. 2018) did not demonstrate any significant change in interval cancers after one round of screening. In this study, the additional cancers detected were small, node-negative cancers and molecular

subtypes recognised to have a good prognosis. These findings are supported by recent studies (Conant et al. 2019, Hovda et al. 2020).

If DBT is to be used as an adjunct to DM, there is also the associated significant increase in the mean reading time (Sechopoulos, Teuwen, and Mann 2020, Tagliafico et al. 2017). Therefore, further research is required to evaluate the features of interval cancers and DBT, combined with results from consecutive screening episodes, to fully comprehend the potential benefits and harms of implementing DBT in screening programs (Hovda et al. 2020). The UK PROSPECTS Trial is a multi-centre prospective study (Michell and Batohi 2018) aiming to address these questions. The Tomosynthesis Mammographic Imaging Screening Trial (TMIST) is a large randomised multicentric study aiming to assess whether DBT combined with 2D-mammography is more effectual in decreasing the incidence of advanced breast cancer (National Cancer Institute 2020).

2.8.2 Contrast Enhanced Spectral Mammography (CESM)

In dual-energy contrast-enhanced mammography low energy and high-energy images are acquired after the administration of a contrast agent. These images are used to construct a recombined image. James and Tennant (2018) advocate that CESM should be considered as a first-line test in patients presenting symptomatically with a clinically palpable abnormality replacing conventional FFDM. Conversely, a recent systematic review and meta-analysis (prospective studies only) (Suter et al. 2020) reported that CESM demonstrated a sensitivity of 85% and a specificity of 77%, and with a 20% false-negative rate is currently considered suboptimal as a first-line diagnostic test. The authors state that CESM might be used as a second-line

investigation in situations when MRI is contraindicated. There is ongoing research to evaluate the role of CESM in screening high and medium-risk patients.

2.8.3 Computer-Aided Detection (CAD) /Artificial Intelligence (AI)

Mammography Computer-Aided Detection (CAD) systems are fundamentally based on highly complex pattern recognition. They are designed to aid reader perception by marking areas for the interpreting reporter to reconsider as a potential abnormality. However, historically, while achieving high sensitivities, CAD systems had low specificities, and the benefits of using this technology remain a topic of debate. A retrospective review was undertaken by Lehman et al. (2015) on CAD in screening mammography and concluded there was no significant effect on cancer detection rates, and the sensitivity was significantly decreased when Radiologists reported with CAD compared to without. Helvie (2007) acknowledged this potential weakness in a clinical application stating that CAD can affect human behaviour or decision-making. This is especially pertinent if readers used the CAD as a first-line tool highlighting potential abnormalities, rather than reviewing the visual prompts after the image analysis. Alberdi et al. (2004) speculated that this might be a result of automation bias (Radiologists vigilance decreased) or characterisation bias (Radiologists defer to CAD instead of relying on their findings). As discussed previously, the recall rate is one of the performance measures utilised in monitoring screening units with standards set to avoid excessive false-positive recalls. A seminal trial (CADET II) (Gilbert et al. 2008) conducted in the UK demonstrated no statistical difference in cancer detection rates when comparing a single reader with CAD and two readers. However, the overall recall rates were higher and significant ($P<0.001$) compared to double reading.

2.8.3.1 Machine Learning

New cognitive technologies, which are advancing rapidly, present the possibility of substantially improving CAD not only for radiology, but for images from pathology laboratories, and combining them with supplementary diagnostic data. The principal technology is deep neural networks (deep learning) to improve the efficacy of imaging-based diagnosis. Machine learning is a form of artificial intelligence (AI) in which computer algorithms can learn and improve directly from the data utilising artificial neural networks (ANN). Deep learning is the type of machine learning that uses multiple ANNs, and the UK breast screening programme provides an ideal database with known pathological outcomes in cases that have undergone a Needle Core Biopsy. There are many potential benefits of assimilating AI into clinical practice, and these are discussed further in Chapter 9.

2.9 Summary

This chapter has described the complexity of mammographic interpretation, the nature of breast cancer and the limitations of current imaging techniques in detection of the disease. The complexity of performance measures within a breast screening setting and the variability in service outcomes has been critically reviewed. The international variance in breast screening systems was highlighted.

An essential factor identified was that new technologies might be effective in detecting more cancers and subsequently result in more recalls to assessment. Therefore, the process of arbitration becomes paramount in reducing the excess of false positives, as human resources and capacity within assessment clinics are hard-

pressed in some services. Although technological advancements have been made in the equipment (FFDM) and techniques, (DBT, CESM) currently, the interpretation of the images is still crucially dependent on individual human decision-making skills. The interpretive performance of mammography is variable, and human decisions are prone to error. The next chapter will review the theory and the complexities of human decision-making.

Chapter 3. Human Decision-Making: A Review of the Literature

The last chapter gave an overview of the breast screening program exploring the challenges in reporting screening mammography. It concluded that the interpretation of the images is dependent on individual human decision-making skills, which are prone to error. Pearson (2013) proposes that the challenges associated with decision-making are amplified in a healthcare setting as a result of increasing workloads, reduced resources and complex patient presentations. Therefore, to understand the theoretical and practical implications this section reviews the published literature, beginning with a definition, moving on to factors affecting the process, and concluding with the models that underlie decision-making in clinical practice.

3.1 Clinical Diagnosis

Croskerry and Nimmo (2011) affirm that in all medical domains, thinking, reasoning, and clinical decision-making are the essential skills underpinning the process of diagnosis. However, there is a general assumption that these skills are instinctively learnt throughout medical training. The probability of error is higher in the diagnostic radiology setting in which subjective interpretation of images is undertaken. Evidence from several studies (Brady 2017, Khullar et al. 2015, Schiff et al. 2009 and Berlin 2007) implies decision-making is not a reliable process as diagnostic errors are frequent (10-15%) and undervalued. Berlin (2007) reports average daily real-time Radiologist error rates of 3–5% but with a retrospective review of studies, this increase to averages of 30%. Discordance between two reporters is acknowledged in all fields of radiology with a 5.4 % disagreement rate

reported within an Accident Emergency study (Hardy, Snaith, and Scally 2013). Quantifying error is easier to achieve when a solitary person is involved in the image interpretation and final report (as per single third reader arbitration). However, this is more difficult when multiple opinions are included in the process (as per consensus meetings). Bankier et al. (2010) state that even when two readers agree or achieve consensus on an image, this does not necessarily equate to a correct decision. However, of more significant concern is the high false-positive rate of up to 61% cited by Nelson et al. (2016) in screening mammography, which has substantial consequences on patient anxiety and morbidity.

3.2 Defining Clinical Decision-Making

The literature review revealed inconsistencies in the terminology utilised to define decision-making. In healthcare, clinical decision-making was the predominant phrase (Trimble and Hamilton 2016, Crebbin et al. 2013, Banning 2007, Pearson 2013). However, diagnostic reasoning (Elstein, Schwartz, and Schwarz 2002), clinical judgment (Redelmeier et al. 2001), critical reasoning (Baird 2008) and critical thinking (Pieterse, Lawrence, and Friedrich-Nel 2016) are all used relative to decision-making. Within the appraised literature, the terms judgement and decisions were also used interchangeably.

Barrows and Tamblyn (1980: 19) define clinical reasoning as

'*the cognitive process that is necessary to evaluate and manage a patient's medical problem'*.

Croskerry (2003) and Graber et al. (2005) assert that clinical reasoning processes are a factor in diagnostic error and therefore understanding the process and contributory factors are essential in improving diagnostic accuracy. This view is also supported by Stark and Fins (2014) who maintain that although the body of knowledge relative to cognitive errors and medical error is expanding, there is still a deficiency in the evaluation and implementation of strategies to improve critical thinking skills and medical judgement.

Clinical reasoning is a cerebral function that involves judgement under uncertainty and may be facilitated or impeded by the work system. The deficiencies of medical clinical reasoning were emphasised over 70 years ago (Bakwin 1945) but have gained more attention in the last decade.

Thompson and Dowding (2002) express clinical decision-making as a selection between alternatives. In breast screening arbitration, this pertains to whether the arbitrator can perceive a mammographic abnormality identified by one of the readers and consider if the imaging features of the abnormality indicate the presence of malignancy. An abnormality may be present but benign, for example, a cyst, or fibroadenoma. It is the clinical decision-making skill that is required to prevent over-recalling of benign lesions while not erroneously discharging a malignancy. However, as discussed in Chapter two, there may be an overlap in mammographic features common to both, and therefore the judgement becomes intricate. Banning (2007) affirms that clinical decision-making is a complex activity. Therefore, for Allied Health Professionals to undertake independent clinical decisions requires training and education that provides and advances the clinical

acumen and cognitive skills to process complex information and make sound judgements. The results of Pieterse and colleagues (2016) implied that the majority of third-year radiography students lacked critical thinking skills at the level demanded. This stance is supported by literature relating to critical thinking skills within other Allied Health Professionals; occupational therapy and nursing students (Çubukcu 2006, Velde, Wittman, and Vos 2006).

3.3 Factors That Affect Decision-Making

Clinical decision-making is a method which requires a combination of experience, knowledge, awareness, peer support and evidence-based practice to guide the process (Anderson et al. 2013). In a healthcare setting the system encompasses not just the healthcare professionals and patients but includes management structures, organisational policies and procedures, leadership styles, staffing levels, unit sizes, interpersonal communications, and resources available for delivering care (Kahneman 2011., Lipshitz et al. 2001 and Adams, Greiner and Corrigan 2004). Klein (1998: 151) reliably found that in unclear situations expert decision-makers could

'*detect patterns and typicality in a glance and realise that they have seen it* previously', define the situation as typical or atypical, and quickly make effective decisions.

The cores skills of clinical decision-making are summarised in Table 12.

Table 12 Principal Skills Required in Clinical Decision-Making.
(Taken from NHS Education for Scotland (NES) Effective Practitioner n.d.).

• Pattern Recognition:	Learning from experience
• Critical Thinking:	Conceptualise, analyse, and evaluate information from a variety of sources
• Evidence-based approaches:	Integrating clinical expertise with research evidence and best practice guidelines into the decision- making process
• Communication Skills:	Active listening, individual contributions respected
• Teamwork:	Support and advice characterised by trust, respect, and collaboration of the team members to assist in decision-making
• Sharing:	Partaking in giving and receiving feedback on decision-making with peers. Learning from cases
• Reflection:	Consciously analyse decision-making. Critically examine and evaluate outcomes as a process of continual learning to improve service delivery/patient outcomes.

Each element has the potential to impact effective decision-making. In a perfect setting, decisions would be impartial, with prior imaging available, no resource constraints, time pressures, interruptions or diminished vigilance. In clinical practice, this is rarely the reality, and clinical decision-making requires a balance of the factors to make an informed decision.

The literature identifies that clinical decision-making is complex. Therefore, to explore clinical decision-making and the inferences for Radiographers undertaking arbitration, theories of decision-making are concisely reviewed to gain an understanding of the constructs that underpin decision-making in practice.

3.4 Theories of Decision-Making

Over the last 30 years, social and cognitive psychologists have been engaged in categorising barriers and facilitators that influence the effective use of evidence in decision-making (Bell et al. 1988 and Kahneman and Klein 2009). Biases and irrational stratagems in decision-making have been investigated in non-medical fields, particularly the aviation industry with analogies compared to the medical field (Bornstein, Emler, and Chapman 1999, Dolan 1999).

Croskerry and Nimmo (2011) emphasise that difficulties in teaching decision-making relate to there being a minimal agreement regarding the process itself, with a diversity of paradigms. The dual-process theory has transpired to be the consensus model providing a varied and robust approach. The most basic version of the theory suggests that there are two modes of thinking, with distinctive characteristics that provide an understanding of how information is processed and the cognitive components of decision-making. These are summarised in Table 13 (Croskerry 2009).

Table 13 Selected Characteristics of Type 1 and Type 2 Decision-Making Processes (Taken from Croskerry 2009).

Characteristic	**Type 1**	**Type 2**
Reasoning style	**Intuitive** Heuristic Associative Concrete	**Analytical** Normative Deductive Abstract
Awareness	Low	High
Verbal behaviour	None to minimal	Yes
Action	Reflexive and skilled	Deliberate and rule-based
Automaticity	High	Low
Speed	Fast	Slow
Channels	Multiple and parallel	Single and linear
Effort	Minimal	Considerable
Cost	Low	High
Vulnerability to bias	Yes	Less so
Reliability	Low and variable	High and consistent
Errors	Common	Few
Hard-wired	Maybe	No
Scientific rigour	Low	High

Type 1 intuitive reasoning is typified by a gut feeling and is a fast, spontaneous, reflexive approach requiring little effort. Intuition encompasses thinking that may involve several characteristics, for example, biases, heuristics, prejudices, emotion, and lateral thinking. Croskerry (2009) assert that type 1 reasoning is less reproducible than type 2 decision-making and is prone to error. Non-analytical models are interpreted via pattern-recognition and attempt to comprehend clinical reasoning by how individuals' group and classify practices.

Type 2 analytical thinking is at the opposite end of the spectrum (Elliott 2010 and Evans and Stanovich 2013), characterised by a slow, deliberate, rule-based approach that requires individuals to generate mental models. This system requires a considerable effort but is associated with few errors. These models imply that in radiology, a diagnosis is made by combining patient presentations to the exemplar stored in the individual's memory (Norman 2005 and Levitin 2002). Higgs (2008) also describes that experienced clinicians pattern recognition may engage illness scripts, which incorporate detailed knowledge of the disease such as risk factors, pathophysiology, signs and symptoms.

Heuristics also termed decisional shortcuts or cognitive strategies are involuntary and allow individuals to expedite judgements and decisions. Although heuristics can enable decision-making, they may also introduce errors, particularly in cases presenting with atypical symptoms (Kahneman 2011 and Lipshitz et al. 2001). Failure of a heuristic strategy can lead to cognitive biases (predictable errors in judgment from reliance on heuristics). Kahneman (2011) advocates that continued learning in a traditional setting creates efficient heuristics, while ambiguous and unstable

environments are a primary cause for a heuristic to fail. Dawson and Arkes (1987) contended that heuristics and cognitive biases could have adverse effects on the judgement of probability and synthesis of information which is considered fundamental proficiencies in medicine. There are multiple heuristics and biases associated with clinical decision-making; a summary of those relevant to a breast imaging setting are defined in Table 14.

Table 14 Heuristics and Biases That Can Affect Clinical Reasoning (Taken from Bornstein and Emler 2001, Stiegler and Ruskin 2012)

Heuristic or Bias	Definition
Anchoring Bias	Prematurely accepting prominent features/diagnosis based on the initial impression. No adjustment made even when further information becomes available.
Affective bias	Numerous behaviours (feelings, biases, emotions) influence our judgment. Convincing yourself that what you want to be true is right, instead of less appealing alternatives
Availability bias refers	A propensity to retrieve examples that are easily recalled, common, or seen recently
Context errors reflect	Misinterpretation of the findings resulting in an erroneous conclusion
Search satisfying	The inclination to prevent further searching once an abnormality has been found. The explanation for why other lesions are missed. Accepting the first response that may explain the findings without considering other explanations
Base rate neglect	Disregard the prevalence of the disease.
Confirmation bias	Pursue information that substantiates the hypothesis being tested
Hindsight bias	Overestimate the ability to predict an outcome although there is no objective basis for predicting it.
Overconfidence	Individuals overestimating their knowledge, abilities, and performance

3.5 Models of Clinical Reasoning

Several theories exist as to the relationship between Type 1 and Type 2 dual processing. Brekhus (2015) interpret the dual processes theory as a system in which decisions are principally made using one method or the other and that researchers are merely required to distinguish what actions (questions) are associated with each process. Moore (2017) entitles this system where one process at a time determines the outcome as the "either/or" model. Bonikowski (2016) and Friedman (2016) have described particular research methods as naturally initiating Type 1 processing, for example, surveys and visual cues, while in-depth interviews tend to activate Type 2 processing.

Others, (Vila-Henninger 2015) endorse a "default interventionist" system, in which Type 1 processes are the default responses used in the majority of daily decision-making and are responsible for our moral decisions, but this can be superseded by Type 2 processing (Evans and Stanovich 2013). Type 1 processing frequently delivers correct answers (Kahneman 2011). Type 2 overriding may be considered to produce enhanced decision-making as analytical reasoning is employed, but this does not necessarily ensure a correct decision. The same applies to Type 1 processing superseding Type 2 processing which can result in irrational judgements. Croskerry et al. (2013) advocate that intervention by Type 2 processing is liable to occur with inexperience, lack of knowledge or in an attempt to modify a recognised bias. Kahneman and Klein (2009) and Stanovich (2009) affirm that the processes function best in specific clinical practice settings. Type 1 performs optimally in consistent and specific settings but is suboptimal in indecisive cases. Type 2 processing is

considered supreme in hassle-free, methodical settings. This theory suggests that third reader arbitration and consensus are associated with Type 2 processing where there is the uncertainty of an abnormality or its clinical significance. Elements such as fatigue, distractions, and interruptions may influence an individual's working recollection (memory) and the ability to acknowledge that a Type 1 decision requires re-evaluation (Croskerry 2009). The default interventionist model supports the stance of two unique processing systems; however, they work more collectively than previous theories have implied.

3.6 Dual Process Theory and Diagnosis

More recently, theorists (Evans and Stanovich 2013, Kruglanski and Gigerenzer 2011) have criticised the "either/or" model. Associating Type 1 processes with bias and Type 2 processes with logical deduction was deemed too simplistic, and a misconception as both methods result in correct responses on some occasions and incorrect on others. This theory is supported by evidence that some judgments, for example, simple logical arguments (traditionally equated with Type 2 processing) are achieved easily (Bago and De Neys 2017, Trippas et al. 2016), suggesting in some cases, Type 1 processing is used. Similarly, there is evidence to support that on occasion persuasion judgments may necessitate extra time and energy (Handley and Trippas 2015), and can be dependent upon various mediating factors (Wiswede et al. 2013). Many cognitive and social psychologists, propose that logical and belief-based processing are activated concurrently, and it is not straightforward to differentiate by separable domains (Handley and Trippas 2015, Pennycook, Fugelsang, and Koehler 2015). Data from these studies support a model of parallel processing in

which the complexity of the task governs the speed and accuracy of the response. In circumstances where Type 1 and Type 2 processes unite with the same outcome (i.e. no conflict), rapid decisions are made with high accuracy. In circumstances when the processes conflict, there is the potential for the methods to affect each other.

Croskerry and Nimmo (2011) applied the dual-process model to real-life diagnostic clinical decision-making. This is portrayed in Figure 4.

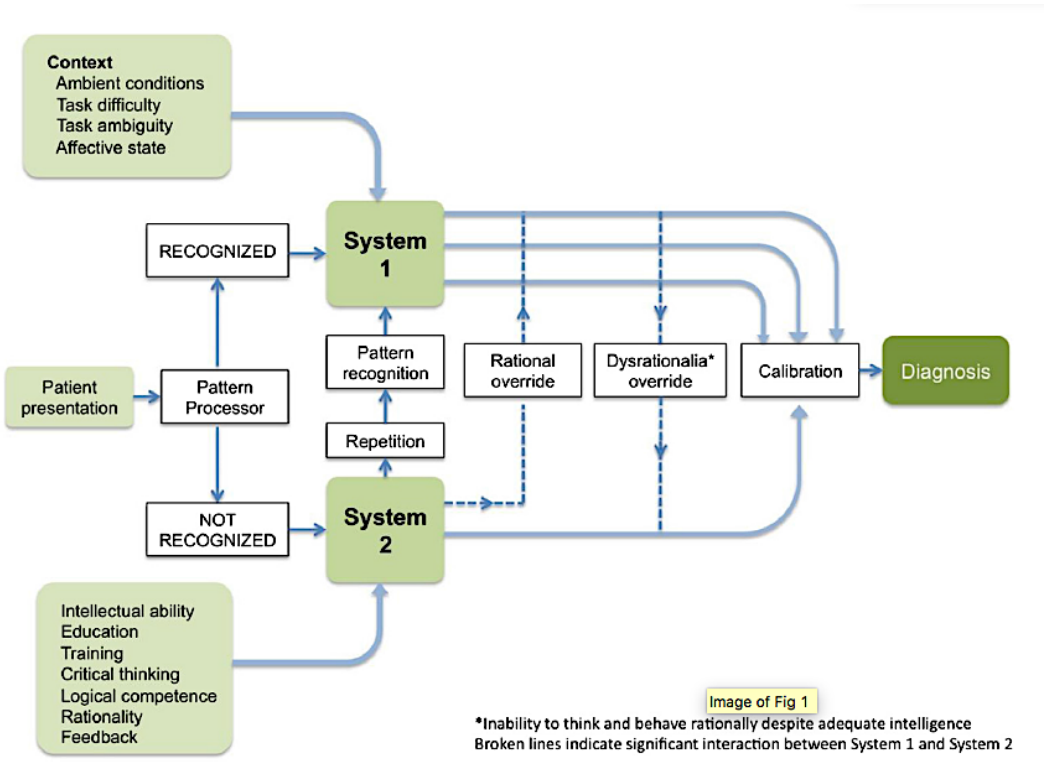


Figure 4 Diagnostic Scheme Based on the Dual-Process Theory.
(Taken from Croskerry and Nimmo 2011).

The model depicts a symptomatic setting in which a patient presents with signs and symptoms of a disease. In a breast screening context, although the majority of women are asymptomatic, there is an opportunity to determine if there are relevant

symptoms, for example, a lump. If the mammographic signs are recognised, there is a high probability that intuition (Type 1 processing) is initiated and a quick decision is made. Conversely, if the mammographic pattern is not familiar, the analytical mode (Type 2 processing) is employed, defined by a slower, deliberate effort to reach a decision. Several features characterise the model:

1. Revisiting the analytical mode will ultimately lead to pattern recognition, engaging intuition. This is the development occurring as expertise is acquired.

2. Logical thinking can supersede intuition if the first impression is considered wrong or requires further contemplation.

3. Intuition can override the analytic mode. Intuition predominates, resulting in an unfounded decision.

4. In the diagram, the blue arrow line denotes swapping between the two processes, representing a dynamic model to produce a sensible decision.

5. The intuitive mode is the default, as it avoids cognitive effort.

In the medical model of dual-process theory, the calibration phase relates to when a patient is reassessed. In practice, this refers to the diagnosis being correct, partially correct or wrong and the patient outcome either staying the same, improving or deteriorating. However, in a breast screening setting the result is only known to be correct if the woman is recalled and confirmed to have a cancer diagnosis or returns for screening in three years with a normal screen. Incorrect diagnosis is likely to

present symptomatically as an interval cancer or will be portrayed as a progression on the subsequent screening mammogram. Therefore, the calibration opportunity for the clinician to switch between intuitive and analytical processes is not applicable, as there is no concurrent patient monitoring.

Length of experience is identified as a contributory factor, with novices predominantly using the analytic mode, while experienced clinicians mainly use the intuitive mode. From a clinical perspective, this is significant for third reader arbitrators as Croskerry and Nimmo (2011) assert that the majority of biases and heuristics are a result of intuition and where many of our thinking failures are derived. In a breast consensus meeting, the dual-process theory implies that the analytical mode offers development as expertise is acquired, and potentially providing safer decision-making.

3.7 Strategies for Reducing Error in Clinical Reasoning

Several strategies have been suggested that improve intuitive performance (Hanoch and Wallin 2003). The elements pertinent to a breast screening environment are summarised in Table 15. However, it is recognised that many of these strategies are directed to inexperienced or novice clinicians.

Table 15 Strategies for Improving Intuitive Performance (Adapted from Hogarth 2010).

Optimise decision-making environment	Optimal decision-making is made in high-quality environments. They should be ergonomically designed and avert interruption and distractions. They should afford the opportunity for expert tutoring and mentoring, providing clinical practice development of domain-specific skills.
Improve feedback	Feedback should be impartial, pertinent, accurate, unequivocal, dependable and delivered promptly. Specify the attributes of the skill that is under emphasis.
Impose circuit breakers	Educate individuals to recognise intuitive actions and alleviate biases. Encourage reflective practice.
Accept conflict in choice	Accept uncertainty and disagreements in any decision. Recognise emotive versus controlled conflict.
Make scientific method intuitive	Promote cognitive forcing functions to prevent biases, rather than trusting intuitions.

Several studies (Coderre et al. 2003, Mamede, Schmidt, and Penaforte 2008, Mamede et al. 2010b) propose that it is a failing of clinical reasoning rather than a deficit of knowledge that cause cognitive diagnostic errors. Thammasitboon and Cutrer (2013) summarise the cognitive solutions to improving diagnostic decisions into the three approaches in Table 16.

Table 16 Strategies for Improving Diagnostic Decisions (Taken from Thammasitboon and Cutrer 2013)

Increase clinical expertise	Individuals identify gaps in their knowledge and skills. Feedback on performance in real life and test case scenarios. Education on the science of decision-making and applying to clinical practice. Engage in continued professional development/CME and competency-based certification.
Avoid cognitive processing errors	Use pattern recognition, slowing down/time-out to avoid faulty intuitive reasoning. Targeted training on errors identified in clinical practice. Improving metacognition promoting reflective practice.
Reduce the cognitive burden	Consult and learn from experts – a second opinion, second reading, a fresh pair of eyes. Use group decision-making. Diagnostic decision support systems – CAD, AI.

3.8 Review of Evidence from Medical Clinical Reasoning Studies

3.8.1 Speed of Diagnosis

The suggestion that Type 1 processing errors can be rectified by taking more time (using Type 2 processing) and using a methodical approach to deliver greater accuracy Evans (2003) has not been conclusively demonstrated in the medical domain. Conversely, Sherbino et al. (2012) reported that rapid diagnosis was accurate. Other investigative studies comparing automatic versus analytical thought demonstrated no difference in accuracy (Ilgen et al. 2011, 2013 and Norman et al. 2014). An experimental study which allowed medical residents to reflect and revise their initial diagnosis (Monteiro, Sandra D et al. 2015) resulted in longer processing times and were significantly less accurate than diagnoses that were not amended.

Evidence from these studies implies that more processing time (deliberation) does not reduce errors, but there is some evidence to support that under time pressure diagnostic accuracy is lower particularly for inexperienced staff (ALQahtani et al. 2016). The literature from the medical studies, therefore, proposes that extra processing time, reflection and identification of biases will have little effect on resolving errors and that knowledge and experience are the major contributory factors of diagnostic performance (Monteiro et al. 2015).

3.8.2 Cognitive Biases

A systematic review (Blumenthal-Barby and Krieger 2015) of biases and heuristics studied in medical decision-making demonstrated that the vast majority of studies pertain to patient decision-making, or joint (patient/clinician) decision-making regarding treatment options. A limited number of studies in the review (n=15)

investigated cognitive bias in diagnostic error, and these were experimental or retrospective reviews of error, with only two relating to actual clinical practice (Graber et al. 2005, and Zwaan et al. 2012). The experimental studies established search satisficing and availability biases, which affected diagnostic accuracy (Berbaum et al. 2013, Hatala, Norman, and Brooks 1999, Schmidt et al. 2014, Mamede et al. 2010b). Contrary to this were two studies where recalled cases enabled accurate diagnosis (Allen et al. 1988, Brooks, Norman, and Allen 1991). However, several of the experimental studies are conducted with students rather than practitioners with years of clinical experience (Norman et al. 2017).

Similarly, some studies have demonstrated that cognitive biases are inclined to diminish with increasing expertise (Weber et al. 1993 and Christensen et al. 1995). Although some researchers encourage education of biases and de-biasing strategies, Norman et al. (2017) assert that the evidence to support this is lacking, with no studies currently demonstrating useful results in practice. Hypothetical scenarios are not synonymous with 'real-world' clinical settings, where decision-making is influenced by technology and teamwork (Patel, Kaufman, and Arocha 2002).

3.8.3 Knowledge-Based Strategies

There are a plethora of studies providing evidence to support that further education and knowledge are related to reductions in error (Custers, Regehr, and Norman 1996, Minda and Smith 2001, Schmidt and Rikers 2007, Norman 2005). However, many of the studies are comparing error rates between junior and experienced practitioners. Schmidt and Mamede (2015) narrative review on teaching clinical reasoning proposes that for novice's knowledge-based approaches where the

underlying process of the disease are appreciated and learning from 'look-alike' diseases demonstrated the most promising results.

Reflective practice has been extensively studied as a method of improving diagnostic accuracy (Mamede et al. 2010a, 2012, Schmidt et al. 2014, Mamede and Schmidt 2004, 2014). Experimental studies have assessed simple versus complex case analysis (Mamede and Schmidt 2004), inducing availability bias to ascertain if reflection would diminish its effect (Mamede et al. 2010a) and intentionally including distraction features (Mamede et al. 2014). Evidence from these studies supports that reflective practice is useful; however, effectiveness is variable dependent on the complexity of the case and experience of the individual. Again, the studies have not been undertaken in a real-life clinical setting. A prospective experimental study by Friedman et al. (2005) which comprised of participants with varying levels of expertise, reported that increasing expertise was associated with greater accuracy and confidence. Expertise comprises of real-life clinical experience supported by feedback on the definitive diagnosis (Kahneman and Klein 2009). The ability to associate past cases with new cases, not only increases the speed of information recall but also the accuracy of a new diagnosis (Brush, Sherbino, and Norman 2017).

3.9 Conclusion

This chapter has provided an overview of the theoretical and practical implications of decision-making. The dual-process theory outlines two types of thinking and the limitations of each approach. A review of the literature substantiates that errors occur in both processing modes. The methods by which clinicians think represents a

valuable part of providing safe healthcare, particularly in the context of correct diagnosis. However, the supposition that training to recognise and reduce biases is effective in clinical reasoning is not supported by the evidence. Likewise, caution directed at slowing down and inducing Type 2 processing had a negligible effect. The precise way in which clinical decisions are made is highly variable, and as yet poorly understood. Studies have identified that environmental and contextual factors influence decision-making. The current study aims to explore decision-making on discrepant breast screening cases and the contributory factors that may influence the process. No literature was identified that directly related to this scenario.

Uncertainty in clinical practice is inevitable. Experience via extensive clinical exposure coupled with regular feedback on patient outcomes provides the knowledge fundamental for improving diagnostic accuracy. A consensus review of discordant breast cases is a collective approach. Therefore, it was felt necessary to examine the literature on group decision-making for the next phase of this research to assess whether this method can enhance the quality of decision-making.

Chapter 4. Review of the Literature on Group Decision-Making

The previous chapter identified the paucity of research regarding the validated efficacy of decision-making theories and models in real-life clinical settings. An overview was also provided of the complexities associated with human decision-making and the associated heuristics and biases. This section reviews the published literature on group decision-making and the team-based nature of consensus processes, critically discussing how practices and team dynamics can influence the outcome and concluding with how algorithms may support the decision-making process in the future.

4.1 Group Decision Making

In group decision-making, a group (classified as two or more people) with subjective experience, knowledge, and attitudes articulate their viewpoint to achieve a consensus decision (Lu et al. 2007 and Montero 2007). The final decision is the responsibility of the group as a whole rather than a specific individual. There are multiple factors entrenched in diagnostic processes and the subsequent decision-making (Tsalatsanis et al. 2015, Donald and Barnard 2012, Trimble and Hamilton 2016 and Balogh et al. 2015). The work system is composed of the task and processes (workflow), the technology, organisational characteristics, the physical environment and the team members. The complexity of systems is confounded by factors such as the structure of the group (size, roles, norms and cohesiveness) the processes (decision-making rules), and factors associated with group dynamics, communication, and group diversity. These are summarised in Figure 5. The confounding factors are critically evaluated.

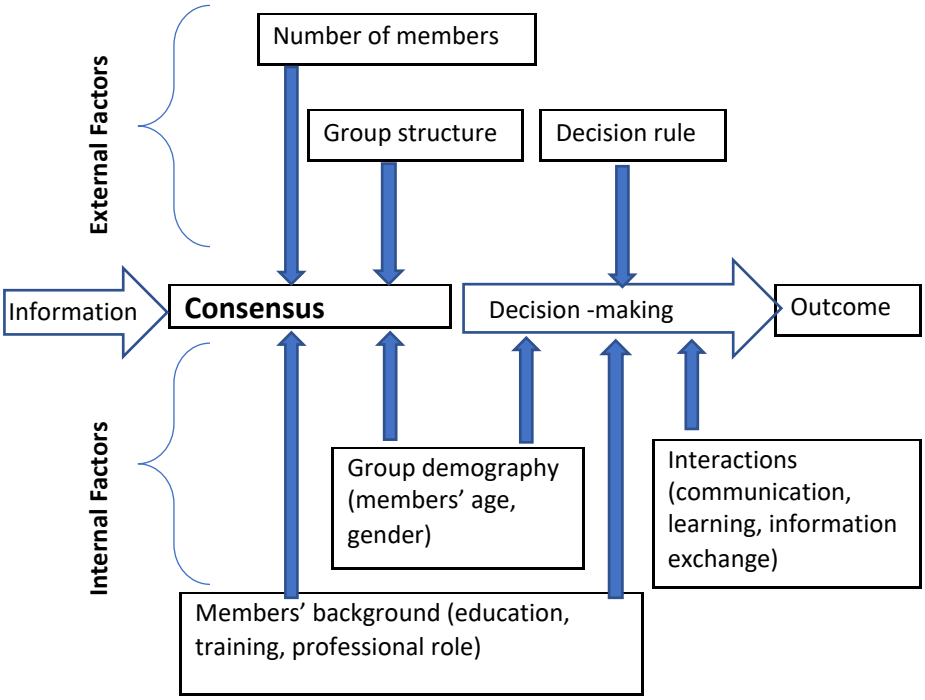


Figure 5 The External and Internal Determinants Influencing Group Decision Making (Taken from Stanek 2013).

4.2 The Decision-Making Process

4.2.1 Decision-Making Rules

There are several methods by which group decisions can be made (Taylor et al. 2013). The standard procedures are detailed in Table 17.

Table 17 Decision-Making Methods (Taken from Taylor et al. (2013).

Unanimity- all members of the group agree on the outcome.
Consensus – through discussion, the group achieves a decision.
Authority – the leader of the group, makes the final decision.
Majority – all group members state their opinion, the majority judgement informs the final decision.
Minority- an individual dominates the group into a decision.

The term consensus suggests a shared endeavour towards an agreement, which is preferably obtained via active collaboration rather than passive compromise (Bankier et al. 2010). Consensus represents a united decision where individual opinions have been considered, although this does not necessarily denote that everyone's voice within the group is given an equal weighting. A key element of consensus is the ability to actively listen to others viewpoints, and a willingness for individuals to change their view following rational argument and persuasion (Rothstein 1987). Differences of opinion are inherent in the subjective area of breast radiology, and a unanimous agreement on a case may be challenging to obtain. Consensus decision-making requires openness, trust, cooperation and respect for team members to concur (Simons and Peterson 2000). A consensus that is easily attained may reflect an environment where individuals do not feel comfortable or confident enough to express their disagreement.

4.3 Organisational Behaviour

Schermerhorn (2012) defines organisational behaviour as

> *'An academic discipline devoted to understanding individual and group behaviour, interpersonal processes, and organisational dynamics.'*

The main principle to understanding organisational behaviour is the situation or context in which the behaviour occurs. Johnson et al. (2016) state that the most substantial contextual influence on organisational behaviour is that of culture. Organisational culture is classified as a learned set of shared values and beliefs within an organisation. The culture of an environment defines the boundaries of a

group and influences the way people feel and interact. Schermerhorn (2012) indicates that in organisations with an authoritarian and hierarchical culture, individuals are reluctant to make individual decisions, and seek approval of others, often exhibiting less initiative. Other cultures are described as competitive with an emphasis on performance results, or innovative with the main focus of generating new ideas and systems of work. How an individual interacts within a particular organisational culture reflects on their confidence and job satisfaction.

4.4 The Diversity of the Team

A vital aspect of any organisation is the individuals within it. Curry et al. (2018) affirm that a positive culture utilises the attributes of all members representing a diverse workforce and embraces respect and inclusiveness. Heterogeneity within a team can be beneficial in that a variety of perspectives, experience and knowledge are constructive. Within teams, there will be individual differences and similarities in how people think, feel and behave. Although these disparities may cause difficulties in working together, they can also confer significant benefits by combining the effects of different skills, approaches and experiences (Roth and Markova 2012).

4.5 Team Working

A review of the literature regarding effective team working in a variety of health care settings revealed a plethora of studies (Nagpal et al. 2010, Schroder et al. 2011 and Jesmin et al. 2012). Despite increased significance placed on improving teamwork, the majority of studies related to a surgical setting, Accident and

Emergency, or nursing teams; with the diagnostic setting not well-studied. Valentine et al. (2014) undertook a systematic review of survey instruments designed to measure teamwork in healthcare settings and concluded that a few studies reported tools that comprehensively captured the team dynamics. The review highlighted that measures of teamwork were either too specific or too generalised and lacked standard psychometric criteria necessary for survey validation. Song et al. (2015) developed a framework for team dynamics within a primary care setting, and although in a different clinical context this was pertinent to the current study, as the three main principles of effective team dynamics were applicable:

(1) team performance

(2) member satisfaction

(3) team adaptation

The conceptual model by Song et al. (2015) hypothesised that for a team to function effectively certain organisational conditions must be present, and without these members do not behave or feel like a team. If supportive conditions are in place, the supposition was that members would report more effective and efficient care from both the patient and professional perspective. Decision-making within breast consensus teams involves professionals of varying roles and expertise, and therefore it is considered that the majority of enabling conditions would be transferable to a breast screening group. Factors deemed relevant were assessed as part of the national surveys, and the results are reported in Chapter 6.

4.5.1 Conditions for Team Effectiveness

Performance in an organisational behaviour context is determined by effectiveness. The performance of an effective team relates to the achievement of tasks regarding quality, quantity and timeliness of results. In breast screening, this would relate to units achieving the NHSBSP standards detailed in Chapter 2. Group decision-making may be deemed inefficient comparative to an individual as the resources in terms of personnel and time are greater. This is particularly pertinent in breast imaging given the national shortage of screen reading personnel. Group decision-making requires discussion, consideration, and coordination. As with any group, it is imperative to have structure and leadership to ensure a productive outcome.

Stability in a team was deemed favourable by Song et al. (2015) as it endorses shared responsibility and facilitates group learning. Cohesive teamwork has been associated with improved patient outcomes, efficiency, quality and professional satisfaction (Grumbach and Bodenheimer 2004 and Weaver et al. 2014). However, cohesiveness is reported to decline as a group size increases. Statistical disagreements regarding the influence of group size and individual's competence on group performance are longstanding (Einhorn, Hogarth, and Klempner 1977). Statistical models report that group performance should improve with an increase in group size, with the most significant effect when the individual members are highly competent. These models adopt a statistical method of combining opinions, often with proportional weighting of the individual's input based on competence (Grofman, Feld, and Owen 1984, Shapley and Grofman 1984). Davis (1992) states that realistic data on group performance reveal that human groups are commonly

less effective and relatively inefficient compared to statistical expectancies. This is a result of the social combination and can be attributed to decreased member motivation or the organising and combining of member contributions.

Disparate views exist regarding the intelligence of a group. Some psychologists propose that the intelligence of the group is the average of the individuals, while others imply it is merely that of the most intellectual colleague, who controls and coordinates the group (Deary 2000, Devine and Philips 2001). Woolley, Aggarwal, and Malone (2015) disagreed with both of these theories stating that group performance was not predicted by the intelligence of the individual members but was a collective intelligence (CI). It was proposed that group intelligence was not associated with accumulated knowledge but was governed by the ability of individuals to interpret colleague's emotions; termed social perceptiveness. A Collective Intelligence (CI) study undertaken by Wolf et al. (2015) reported that improvements in true positives and decreases in false positives levelled off around a group size of nine. However, it is essential to note that this study aggregated individual performance data based on three CI rules (quorum, weighted quorum and majority rule) and this was not a face to face interaction and discussion between the Radiologists. Interestingly, Sorkin et al. (2001) report that group efficiency decreased as group size increased and concluded that this is a result of ineffective group function, which may be attributed to complexities in individual interactions, individual effort or difficulties in combining opinions.

Highly cohesive teams may also be more susceptible to the phenomenon of groupthink resulting in performance-reducing effects (Moorhead 1982). Groupthink

is a vital factor which occurs when the desire for agreement or compliance may produce incorrect or unsound decision-making outcomes. By repressing opposing opinions to minimise conflict, individuals change their judgment to what they

"*believe others want to hear"* (Bankier et al. 2010: 16).

The outcome is a 'pseudo' decision made without critical evaluation of different opinions (Schermerhorn 2012), resulting in a consensus that may not be judicious. The phenomenon of group polarisation represents an inclination for some groups to make riskier decisions than they would individually. As a case is discussed, individuals become less cautious, and in strongly cohesive groups, people may feel pressured to conform (Schein 2010).

4.5.2 Supportive Processes

Song et al. (2015) describe three supportive processes which reinforce teamwork:

1. Accountability
2. Decision-making and conflict resolution
3. Communication and sharing of information

Schermerhorn et al. (2012) stipulate that just as an individual is responsible for their performance, a team is also collectively responsible. However, Walters et al. (2016) report that one of the main shortcomings of group decision-making can be a lack of accountability. Comparative to an individual who is responsible for their work and repercussion of their actions, group decision-making can create a diffusion of

responsibility. As there is no single ownership, it is easier for individuals to negate personal responsibility for incorrect decisions.

Group decision-making is purported to be advantageous to individual decision-making as it provides the opportunity for collaboration and information sharing between individuals with varying levels of experience and knowledge. However, the merit of the team is dependent upon the members working collaboratively (Baker, Day, and Salas 2006). The supposition is that a collective judgement obtained via discussion, questioning and teamwork will result in a more accurate outcome than an individual. A systematic scoping review in 2017 (Hackney et al.), found no evidence to support this hypothesis in a breast screening context. Conversely, in breast screening Blanks et al. (1998) concluded that the Standardised Detection Ratio (SDR) was higher for double reading with third-person arbitration compared to consensus. This applied to both prevalent and incident screens and smaller cancers (<15mm).

Although group decision-making is generally inferred to be positive, multiple studies report the problems associated with this process. Bankier et al. (2010) and Wolf et al. (2015) describe the complexities of dynamics that exist within group discussions where one member is dominant, and individuals with strong influential characteristics can coerce decisions. Effective teamwork is dependent upon adequate procedures to manage conflict in decision-making. A certain amount of disagreement is considered advantageous as this explores cases thoroughly, working through the possible outcomes (Huczynski and Buchanan 2013). However, significant conflict can result in strained relationships and initiate negative team

dynamics. Edmondson and Bohmer (2001) state that effective conflict management is essential so that decisions made can be critiqued in a safe setting. This can be achieved by openly discussing an individual's judgements as part of a learning process (Argyris and Schön 1978).

The final critical process to reinforce teamwork is communication and information exchange, which supports safe and effectual care (Weiss and Davis 1985). A breast screening consensus group is commonly comprised of film readers with differing professional backgrounds (Radiologists and Radiographers) and education (medical and Allied Health Professional) with varying levels of experience. A conscious effort is required to disseminate knowledge between team members, to support colleagues with less confidence to encourage their input into the group discussion (Gardner, Gino, and Staats 2012).

4.5.3 Behaving and Feeling Like a Team

Although some studies indicate that those team members who behave and feel like a team, experience a higher level of team effectiveness (Kozlowski and Ilgen 2006 and Brennan et al. 2013), there can be striking disparities in the perceived quality of teamwork reported by the different professional groups (Manser 2009). Several studies have demonstrated nursing staff expressing lower levels of quality of teamwork comparative to doctors (Flin et al. 2006, Thomas et al. 2003, Fleming et al. 2006, and Huang et al. 2007). These differences were also reported within professional groups at varying levels of experience (trainee doctors stating lower levels than senior colleagues) (Flin et al. 2006, Thomas et al. 2003, Fleming et al. 2006, and Huang et al. 2007). Establishing a culture that fosters intra-and

interprofessional collaboration is essential. Taplin et al. (2015) assert that it is the identification and management of distinct but symbiotic roles that differentiate a team from a group. West and Lyubovnikova (2012) describe teams as real and pseudo. In real teams, there is clarity of roles and responsibilities and trust is exhibited between colleagues. Value and respect for team members' views and roles are manifested (Schroder et al. 2011), with individuals experiencing a sense of belonging. This is supported by Searle and Skinner (2011) who confirm that member satisfaction within an effective team is deemed high if individuals believe their contribution and involvement are valued. Conversely, pseudo teams demonstrate silo working, a lack of clarity in tasks and accountability, and subsequently little trust amid individuals. In a power structure, individuals use social power as influences over one another.

AbuAlRub et al. (2012) propose that trust is acquired and supported via effective leadership, which is associated with evident improvements in patient care, quality and safety. Building a culture of trust and commitment is essential in a breast consensus group to gain the associated improvements in performance, efficiencies, behaviours and collaboration.

4.6 Conclusion

Although the value of teamwork is recognised, a recent systematic review of validated survey instruments of team effectiveness in healthcare states

"*There are no consensus strategies to help healthcare organisations achieve optimal teamwork"* (Kash et al. 2018).

Furthermore, Kash et al. (2018) conclude that future research is required across a range of healthcare fields to determine if there are characteristics (skills and behaviours) associated with higher-performing teams (outside of composition) that would facilitate improvements in productivity, effectiveness, and quality. A limited number of studies in their review had patient outcomes as a fundamental component within the survey tools. This is important if improvements in team effectiveness are to translate to improvements in patient care.

This literature review highlighted that the dynamics of consensus group decision-making are complex. Human decision-making is associated with inherent subjectivity, error and imprecision in the expression of opinions. Only a limited number of studies have utilised CI in medical decision-making. Some studies reported that diagnostic accuracy improved with group decision-making (Wolf et al. 2015, Kurvers et al. 2015, Hautz et al. 2015 and Kattan et al. 2016), whereas other studies found insignificant or adverse effects (Kee, Owen, and Leathem 2004, Christensen et al. 2000). Wolf et al. (2015) propose that CI rules offer several advantages to conventional direct group discussion and interaction. Algorithmic CI rules offer a transparent collective decision circumventing the concept of groupthink and preserving diversity. Also, convening a face to face group consensus may be difficult to facilitate in busy departments with limited staffing resources.

Conversely, CI rules only require an independent review which may also confer valuable time efficiencies. Furthermore, with the advent of AI, several questions for future research are raised in how this new technology may enhance decision-making in breast arbitration cases. This is discussed further in Chapter 9.

Chapter 5. Methodology

The preceding chapters have established the limited body of evidence relating to arbitration practices within breast screening and the complexities of human decision-making. This chapter discusses the rationale underpinning the design and methodology adopted in this research. Ethical considerations are presented, critically reviewing the measures to protect participants from harm. Study quality and rigour are introduced, but as three distinct study phases are undertaken, the detailed methods are described in the respective chapters (6,7 and 8).

5.1 Research Problem

This thesis explores the current variation in reporting and arbitration strategies within breast screening services in England. It correlates findings with performance based on specific criteria from published national service data (KC62 2013/2014 - 2016/2017).

The study sought to explore and explain:

- what factors determined the strategies used
- how services were organised
- the implications of the varying strategies
- what factors affected the implementation of Radiographer arbitration and hence, what was the effect within England of the new PHE arbitration guidance?
- the future role of new technology, in particular, Artificial Intelligence (AI) in this setting.

5.2 Methodological Considerations

There are diverse world views on what constitutes authentic knowledge and relevant subjects to research (Gerrish and Lacey 2012). Lincoln and Guba (1989: 221) define a paradigm as a '*basic belief system that guides the investigation'.* This can be characterised by **ontology** (the assumptions regarding the nature of reality), **epistemology** (beliefs on how you know something or might discover knowledge) and **methodology** (the tools and techniques used to conduct the research). Together these characteristics create a comprehensive understanding of our perception relative to culture, how we interpret knowledge and the subsequent methodological approach utilised (Crotty 1998).

Morse et al. (2001) maintain that explicit philosophical assumptions support different methodologies, and that consistency between the philosophical basis and methods produce more valid results. Underpinning the positivist paradigm (scientific approach) is the belief that there is a measurable reality, and thus a quantitative approach is utilised (Keele 2011). The naturalistic paradigm seeks to gain an understanding of people (actions, decisions, beliefs, values) in their social world, aiming to examine the phenomenon from the perspective of those experiencing it. Within this paradigm, a qualitative approach is employed (Keele 2011). There is a general view that mixed methods research is relatively new, but De Lisle (2011) describe studies dating back to the 1920s and '30s. However, mixed methods have progressively developed (Creswell and Plano Clark 2011), officially recognised within the last twenty years (Teddlie and Tashakkori 2009).

Denscombe (2008) deems mixed methods as the third paradigm, but Stockman

(2015) argues that this is unhelpful to surmounting the lasting prejudices as it propagates a paradigm debate. While there is continuing debate about which worldview(s) mixed methods research (MMR) associates with, pragmatism is commonly accepted (Creswell and Plano Clark 2011 and Tashakkori and Teddlie 2010). This thesis utilised a mixed-methods approach to map current arbitration practices within England via a national survey. However, to understand why there is variance in practice and what can be learnt from higher-performing units, qualitative interviews were undertaken. Pragmatism aligns with health service research as prominence is placed on practice and interactions within specific environments (Pluye and Hong 2014).

The principal of a mixed-methods approach is that combining and integrating quantitative and qualitative approaches enhances the understanding of research problems relative to a single approach, allowing flexibility to optimise quality (Creswell and Plano Clark 2011). This concept is summarised by (Baars, 1980: 15)

"without naturalistic facts, experimental work may become narrow and blind: but without experimental research, the naturalistic approach runs the danger of being shallow and uncertain".

Four main types of mixed-methods designs are described in the literature; explanatory, exploratory, embedded and triangulation (Mertens 2005, Tashakkori and Teddlie 2010). Mixed-methods research is advantageous for understanding complex health settings in which individuals and environmental factors influence behaviour, policies and systems of work (Ivankova and Kawamura 2010). Utilising

mixed-methods allows researchers to explore multiple perspectives, outline trends, evaluate and triangulate findings, and assess processes and outcomes (Creswell and Plano Clark 2011). Blandford (2013) describe four different methods of triangulation (Table 18). In the current study, methodological triangulation was used (highlighted) as the results from the interviews were used to strengthen and, or explain the results from the surveys and KC62 data (Morgan 1998). Although triangulation can lead to convergent results or complementary results, it can also highlight divergent findings which require further exploration (Erzberger and Kelle 2003).

Table 18. The Four Triangulation Methods and Associated Descriptors

(Taken from Blandford 2013)

Triangulation Methods	**Description**
Data triangulation	Comparison of data from varying sources; may support generalisability of findings.
Investigator triangulation	Data is collected and analysed by various researchers.
Theory triangulation	Employing different theoretical frameworks.
Methodological triangulation (used in this study)	***Using various data gathering techniques to corroborate findings.***

Chapter 1 described how third reader arbitration and consensus teams may be entrenched within complex organisational structures, and therefore a mixed-methods study was deemed appropriate to establish 'what works, in what context, for whom, and to what effect?'.

5.3 Research Design: A Mixed-Methods Approach

A mixed methods research design has a philosophical framework that influences the collation and data analysis in several phases of the research process. This approach allows the researcher flexibility in emphasising the quantitative or qualitative component or equal priority given to both parts (Molina-Azorin 2016).

The specific questions that the study sought to answer were:

1. Is there variation in the approaches and processes used in decision-making within the 80 breast screening units within England, and how has this developed over time? What is the future role of new technology?

2. What are the potential barriers and facilitators of different decision-making processes (along the arbitration spectrum) in breast screening?

3. What are the implications (time to report/clinical resources & skill-mix/ perceived benefits) of the different strategies; and does time/resources invested in reviewing concordant recalls result in any significant reductions in recall rates?

4. What are the implications within England of the new PHE arbitration guidance?

5. What can be learnt from decision-making in higher and lower performance units to inform the future efficient use of arbitration processes in breast screening?

Within this study, the collection of data was undertaken in phases (sequential design) with the primary administration of the surveys to obtain national quantitative and descriptive data of practice from a Director and breast screening

reporter perspective. Second, quantitative unit performance data (KC62) was collected and used to stratify upper- and lower-unit performance. A secondary purpose of using MMR was connected integration, with survey responses and performance data used to define a sampling frame for the interview phase (Curry and Nunez-Smith 2015). Qualitative interviews also allowed the investigation of staff opinions on how the arbitration guidance may be improved together with barriers and facilitators of Radiographer arbitration. Interviews were also critical for the explanation and expansion of particular findings emergent from the quantitative data, for example, the organisational variance in reporting and arbitration practices, and interesting comments provided in the free text. Each method had equal weighting. Figure 6 demonstrates the phases of the sequential study design.

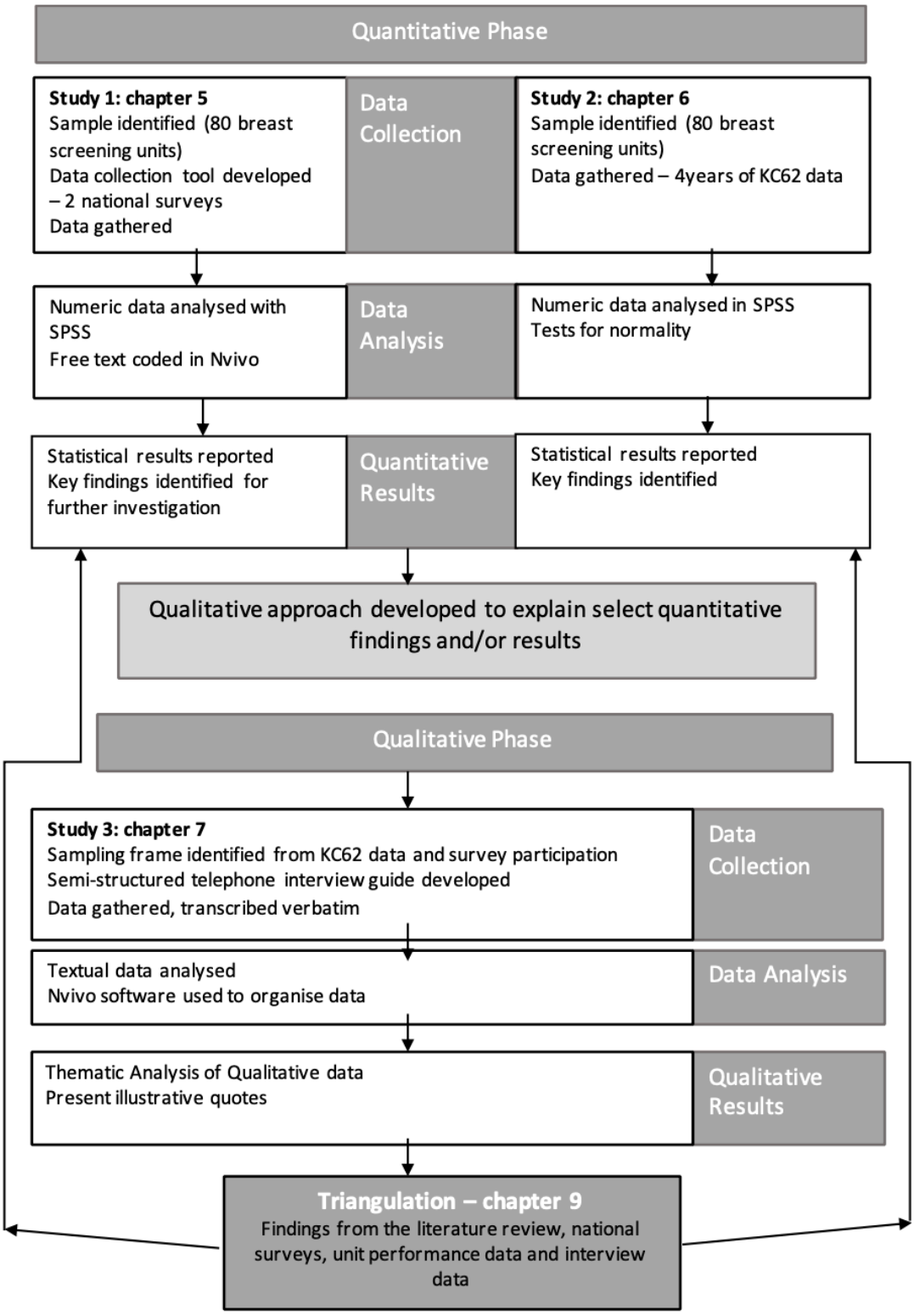


Figure 6 Demonstrating the Sequential Explanatory Study Design and Mapping to the Thesis Structure.
(Adapted from Curry and Nunez-Smith (2015)

Mixed methods research enables triangulation of one set of results with another to enrich the knowledge gained and increases the validity of inferences (Creswell and Plano Clark 2007). Within this study triangulation of the literature review was undertaken with unit performance data, survey results and interview data to establish if there are common characteristics within higher performing units relative to the processes used, the skill mix, and unit size. However, there are disadvantages to MMR, which are detailed in Table 19.

Table 19 Advantages and Disadvantages of MMR (Denscombe 2008)

Advantages	Disadvantages
A comprehensive explanation of the research topic by merging insights from multiple perspectives. A combination of qualitative and quantitative methods enable exploratory and explanatory research; results tend to offer a broader range of questions relating to 'what' 'why' 'how' 'when' and 'who'.	Labour intensive and may be costlier. Data collection and analysis may be more time-consuming.
MMR emphasises the integration of data, how the data compare and contrast and the benefits/complement of multiple sources	A requirement for the researcher to acquire and employ skills in multiple methods (qualitative and quantitative approaches). More demanding on a novice researcher.
MMR embraces triangulation – validate data via cross verification from multiple sources to demonstrate consistency Practical, problem-driven method of research	Complex methodology Pragmatism may be misconstrued. Findings from different methods may not substantiate each other. Further research may be required to explain the disparity.

5.4 Ethical Considerations

Researchers encounter ethical challenges at all stages of a study, starting from the initial proposal to final reportage. Governance structures are in place to primarily protect research participants from harm and to ensure that they are always treated

with respect (Hope, Savulescu, and Hendrick 2008). However, researchers must be conversant with several considerations to include informed consent, anonymity, confidentiality, researchers' potential influence on the participants and contrariwise (Sanjari et al. 2014). The key ethical issues for this study were anonymity and confidentiality.

5.4.1 Impact on Participants

In qualitative studies, the relationship formed between the researcher and participants is paramount to achieving honest and open interactions, preventing misrepresentations and subsequently acquiring useful quality data (Richards et al. 2002). The researcher must attempt to minimise intrusion on participants in clinical practice (Coombs and Ersser 2004), and therefore the surveys were open for six weeks to facilitate time for completion. Telephone interviews were carefully planned to accommodate available days/times within the participant's commitments. The study was conducted in accordance with Coventry University research ethics and governance committee; the principal ethical considerations are detailed below.

5.4.2 Informed Consent

Informed consent is an integral part of ethics in research. Informed consent represents the agreement given by a research participant to take part in a clinical research study, and for validity, this must be: - informed, voluntarily given and the participant competent to consent. The participant information sheet specified in advance the data that would be collected and clarified the nature and objectives of the study. The researchers identify, the participants' potential role and information

on how results would be utilised and published were also detailed. The first page of the online questionnaire also provided this background information. Before commencing the questionnaire, participants were required to answer a mandatory question “Do you agree to take part in this study”, Yes or No. If they agreed to take part, they were automatically directed to the questionnaire. If they declined to take part, they were automatically directed to the end, therefore being unable to view or complete the questionnaire.

The online surveys had a final section which asked participants if they would be willing to partake in further research associated with the project in the form of a semi-structured telephone interview. If willing, they provided a contact e-mail address. Ethically this had to be kept separate so that it remained anonymous. Individuals were e-mailed the study participant information sheet and a consent form. After first contacting the potential participants, they had two weeks to decide if they wanted to partake in the study. Radiographers and Radiologists were requested to return the signed consent form to indicate their willingness to participate.

Consent was re-affirmed verbally just before each interview commenced. Interviews were audio-recorded with the approval of the respondents. Participants were able to withdraw from the study at any time before and during the interviews. The Participant Information Sheet explained that after this time, it would not be possible to extract data from the study as concurrent data analysis would be performed as interviews were completed.

5.4.3 Anonymity

The concept of confidentiality is reinforced by the principle of respect for autonomy and denotes that identifiable information collected about individuals during the research process will not be disclosed (Bryman 2015). Anonymity is one method by which confidentiality is accomplished. Discussing departmental issues can be difficult, and consequently, healthcare professionals may have been reluctant to talk openly. Therefore, throughout the process, participants were assured that the information provided would remain confidential. To protect the identity of the staff and the unit, each participant was allocated an anonymous study-specific code. This safeguarded participant identities in survey data, interview transcripts and subsequent research dissemination. Participants were requested not to mention staff names in their interviews but were reassured that inadvertent disclosures would be removed from the transcripts.

5.4.4 Confidentiality and Data Protection

Ethical guidelines are explicit that confidentiality is a principal element of social research and that participants should be made aware of who will have access to their data, as well as being informed about the procedures for anonymisation (Oliver 2010). Study data was handled in accordance with the UK policy framework for health and social care research (*UK Policy Framework for Health and Social Care Research - Health Research Authority* 2017) and Coventry University safeguarding data policy. All identifiable information (electronic data) retrieved relative to breast care units, performance data, individual participant's (consent forms) and interviewees (interview transcripts) were stored electronically on a secure (Coventry university password protected) server (student one drive). Digital audio recordings

of interviews were uploaded daily to encrypted data sticks and the secure University server. All data is scheduled to be deleted from the researcher's data stick and University 'one' drive three years after the PhD has been completed.

5.4.5 Disclosure

The literature states that researchers must consider situations that may require confidentially to be broken, and there is an obligation to explicitly state this in the consent process (Ritchie, Jane and Lewis 2003). The risk of disclosure was considered extremely low within this study as there was no direct patient contact, observation of clinical practice or intervention. Surveys and interviews covered the professional practices of existing NHS Radiographers, Radiologists and Breast Clinicians and did not relate to personal or sensitive information. The provision was made to discuss any potential bad practice described with research supervisors. Before commencing the telephone interviews, participants were informed that in this scenario, they would be notified of any action taken. No disclosures occurred during the study.

5.4.6 Risk of Harm

The researcher had undertaken Good Clinical Practice (GCP) training. A primary value of ethical research is to minimise the risk of harm to participants (Hope, Savulescu, and Hendrick 2008). The principles of beneficence and non-maleficence are adhered to (Polit and Beck 2016, Offredy and Vickers 2010, Gerrish and Lathlean 2015) within this study as no significant harm was foreseen to the NHS staff from taking part in the research surveys or interviews. The interviews were undertaken by telephone, requiring no fieldwork, and therefore no harm was foreseen to the

researcher. The participants openly documented in the surveys or willingly talked about their practice and experience (positive and negative) in the interviews. Hence, it was considered that the participants were reassured; there was no harm by partaking.

5.4.7 Ethical Approval

Ethical approval for the study was granted by Coventry University Research Ethics and Governance Committee stage 1 *(reference: P45921)* on (06th February 2017) and stage 2 *(reference: P50587)* on (26th May 2017). Following the Health Research Authority (HRA) guidance (2016) studies led from England, involving the NHS in England, should now obtain HRA Approval via the Integrated Research Application System (IRAS). HRA approval (IRAS:228030) was received on 31st July 2017 (Appendix 5).

5.5 Study Quality and Rigour

Four main factors reinforce quality and rigour in research studies. The conventional standards of quality and the criteria used for evaluating qualitative and quantitative studies are summarised by O Cathain (In Tashakkori and Teddlie 2010) in Table 20. Examples of the strategies used to ensure rigour in this study are included.

Table 20 Conventional Standards of Quality and Appraisal Criteria.
Sources: Adapted from (Ritchie et al. 2014, Polit and Beck 2010)

Standard	Qualitative Appraisal Criteria	Quantitative Appraisal Criteria	Examples of strategies to ensure rigour in this study
Veracity	Credibility – The degree to which the results plausibly explain the subject under exploration	Internal Validity – The extent to which the findings signify an accurate indication of a causative relationship between variables	▸ Representativeness: Surveys distributed to all 80 breast screening units in England The sample of 18 professionals working within the NHSBSP willing to share their experiences of arbitration/consensus enabled in-depth clarification. The interview sample was taken from a pre-determined framework and included units from a wide geographical area. Utilising verbatim extracts from the free text survey comments and interviews allows the reader to judge if definitive themes are authentic Impartial representation of varying experiences allowing comparisons and contrast between different units and practitioners Interview protocol tested using a pilot interview Data triangulation from survey comments and interviews
Consistency	Dependability – The extent to which the changing context and circumstances are documented and defined. Relies on documentation of transparent decisions and the 'trustworthiness' by which the research has been conducted.	Reliability – repeatability of findings. The dependability of the analysis. Explanation and justification of biases	▸ Achieving auditability: Transparent description of the study from inception, justification of methodology and reporting of findings. Recording decisions and the rationale for them, documenting challenges to sustain consistency between the study's aim, design and methods. Trustworthiness was validated by subjecting the study findings to researcher peer review. A sample of interviews coded by a member of the supervisory team to confirm reliability. Cronbach's α used to determine the internal consistency of the survey.
Applicability	Transferability – the degree to which the results relate to other populations/settings	Generalisability (external validity) – The extent to which results are reliable outside the study population or in other settings	▸ Application to other contexts: Used purposive sampling techniques Provided a rich detail of the study context and phenomenon, including the inclusion/exclusion of participants Potential for findings to transfer to other team settings with inter-professional skill mix and hierarchical structures. Mixed Methods Research with data triangulation
Neutrality	Confirmability –the extent to which the participants rather than the researcher influences the findings	Objectivity – impartiality by the researcher	▸ Reflexivity Reflective journal summarising the researchers understanding of the data acquired and documentation of decisions documented. Methodological triangulation Two of the interviewees were known to the researcher. The researcher remained formal and followed the standard procedural guidelines to maintain rigour.

Demonstrating rigour in mixed methods research is multifaceted and a subject of debate. Tashakkori and Teddlie (2010) describe the evidence on mixed methods quality as inconsistent; regarding the terms used, the concepts that should be evaluated, and data collection/analysis. Several authors emphasise the critical issues of assessing interpretation and integration and the requirement to defend mixed methodology (Wisdom et al. 2012, Curry and Nunez-Smith 2015). The predominant statement is that transparency should be evident in the explanation of the research process so that readers can appraise the quality (Bryman, Becker, and Sempik 2008, Wisdom et al. 2012). The standards of veracity, consistency, applicability and neutrality are discussed in the subsequent individual chapters.

O'Cathain et al. (2008) devised the Good Reporting of A Mixed Methods Study (GRAMMS) guidelines which were considered pertinent to this study as they specifically apply to mixed methods, assessing quantitative and qualitative methods within the design. Table 21 details the pragmatic statements and the sections which have addressed these within this study.

Table 21 The GRAMMS Guidelines.
(Taken from O'cathain et al. 2008)

Statement	Section/Chapter
1) Justifying the rationale for using an MMR approach	Section 5.2 and 5.3
2) Explain the study design, describing the purpose, priority and sequence of methods	Section 5.3
3) Explain sampling, data collection and analysis for each method	Chapter 6,7 & 8
4) Explain at what phase of the study integration occurred, and how it was undertaken	Section 5.3
5) Explain any limitations associated with using qualitative and quantitative methods	Chapter 9
6) Explain insights acquired by integrating methods	Chapter 9

5.6 Chapter Summary

This chapter has discussed the relevance of a mixed-methods approach concerning the current study, describing the precedence, and order of methods. The research questions have been defined and the ethical conduct critically reviewed.

The following three chapters present the three successive stages of the study, detailing sampling, data collection and analysis and the quality and rigour of the individual methods. Chapter six presents the first stage of the study in which national online surveys were undertaken. Chapters seven and eight present the second and third stages of the study which include analysis of unit performance data and critical analysis of the qualitative data collection (telephone interviews) with the theoretical justification of the study sites and participant sampling strategy. Comparison and contrast of data linked with triangulation are discussed in Chapter 9, along with the benefits and limitations of the study design.

Chapter 6. Mapping Current Reporting and Arbitration/Consensus Practice in Breast Screening Units; A National Survey

This chapter discusses the methodological approach of two national surveys, offering a rationale for selection while critically appraising the data collection method. As emphasised in Chapter 2, there is a paucity of research investigating reporting practices and processes to resolve discordant reports in breast screening; hence, the requirement for a survey to map current practice. This chapter provides a systematic analysis of survey responses provided by Directors of Breast Screening Units (Study A) and Breast Screening Reporters (Study B). Descriptive statistics are presented analysing quantitative responses, including number, percentage, mean, and standard deviation (SD) where appropriate. Qualitative analysis of free-text comments is presented to support specific quantitative results.

6.1 Aim and Objectives of the Surveys

6.1.1 Aim

The surveys aimed to explore the development of current reporting and arbitration practices within breast screening units in England, from a Director and breast screen reporter perspective.

6.1.2 Objectives:

1. To identify current reporting and arbitration practices
2. To identify the factors influencing the development of current practice
3. To identify the perceived advantages/disadvantages of the varying strategies, together with the resources required (average number of cases requiring review, time to report/clinical resources & skill-mix)

4. To gather and compare opinions on the PHE guidance on arbitration; identifying barriers and facilitators to implementing Radiographer third reader arbitration/lead of consensus review meetings
5. To identify the impact of the PHE guidance on arbitration

6.2 Methods

A descriptive, cross-sectional survey was deemed appropriate as this provides the ability to collate data from a wide geographical area (Parahoo 2014). Ellis (2016) validates descriptive surveys as a research tool in a situation where little is known about a subject enabling description and comparison of any variance across the units. Additionally, questionnaires offer the researcher the ability to collect quantitative data and qualitative data dependent on the questions and the formatting of how participants are required to respond (Greenhalgh 2014).

6.3 Population and Sampling

Moule, Aveyard and Goodman (2014) stipulate that defining the study population and ensuring a representative sample is essential in survey research. A sample is considered representative if it provides a cross-section of the population that comprises of all relevant factors and variables and provides a balance to the proportions occurring in the overall population. Thus, allowing valid conclusions to be drawn from the research data.

This element of the study comprised of two semi-structured questionnaires; Study A (Director of breast screening units) (Appendix 6) and a complementary survey, Study

B (NHSBSP Breast screening reporting staff of varying professional disciplines) (Appendix 7).

6.3.1 Director of Breast Screening Units in England (Study A)

In Study A, a purposive sampling method of all Directors of breast screening units within England was utilised. Published data (KC62) provides a list of the 80 units. In August 2017, Public Health England (PHE) was used as the primary source for the distribution of the survey via a covering e-mail and electronic letter to Directors at each identified breast screening unit. This approach was chosen to ensure no bias, precise identification of the Directors and ensured current contact e-mail addresses. Also, as an e-mail from a known agency of the Department of Health is less likely to be considered spam and deleted without opening (Edwards 2010). The survey was not extended to the United Kingdom as the arbitration guidance pertains to units within England, and performance data that would be utilised to stratify participants for subsequent telephone interviews would be published Public Health England (KC62) data.

6.3.2 Breast Screen Reporting Film Readers in Units in England (Study B)

In Study B, **s**nowball sampling was utilised as a means of recruiting screen reading Radiographers, Radiologists and Breast Clinicians. The Directors were requested to cascade the link to the film reading survey to relevant staff within their unit (including locum staff). Parahoo (2014) affirm that sampling frames reduce bias, confirming a representative population. The SCoR were contacted, but at present, there is no register of Radiographers undertaking breast screen reading, and therefore the actual number within England is unknown. Thus, a probability

sampling technique could not be utilised. It is acknowledged that the snowball sampling method is not deemed a representative sample for statistical purposes and may introduce bias (Parahoo 2014). However, it is a reputable technique for research involving a population that is difficult to identify or locate.

Participants in both Study A and B were requested to provide the name of their breast screening unit to assist with identification of responses, enable correlation of Director and film reader responses from the same unit and grouping of responses by geographical regions. This information was then anonymised.

6.4 Data Collection Instrument

Survey questionnaires provide the ability to accrue opinions, attitudes, beliefs, and experiences on a large scale (Parahoo 2014). Rattray and Jones (2007) stipulate that it is preferable to use established questionnaires that have verified reliability and validity and therefore allow comparative analysis of study findings. The researcher did not identify any validated questionnaires relating to breast screening arbitration, and therefore a new instrument was constructed.

6.4.1 Types of Survey

A variety of survey modes exist regarding the distribution and completion. Each has intrinsic advantages and disadvantages. These are summarised in Table 22.

Table 22 Advantages and Disadvantages of the Types of Survey and Completion Methods (Adapted from Boynton and Greenhalgh 2004, Jones et al. 2013)

	Advantages	Disadvantages
Distribution method		
Post/Paper	Large-scale study covering a wide geographical area Letters can be deemed a more personal contact No requirement for digital information technology	Cost – associated with printing, packing and postage Delay associated with turnaround time – posting and returning Response rates – generally low Potential for incomplete returns Routing of irrelevant questions potentially more difficult for participants Potential for data inputting and transcription errors
Electronic	Cost-effective Quicker to complete - Online-ease of answering – tick boxes Can be anonymous Ability to make questions compulsory Ability to highlight data entry errors – missing response, too many options selected Ability to indicate participants progress in the survey – may aid completion Re-route inapplicable questions Questionnaire can be saved and restarted at the same point on multiple occasions Provide an invitation e-mail letter and contact details if technical problems encountered Information Technology readily available in hospital environments Data in a format ready for analysis – downloaded to excel, SPSS No transcription errors Quicker turnaround - instantaneous delivery of completed surveys	Dependent on an individual's incentive to partake Potential for participants to misinterpret the question Response rates variable Dependent on IT resources and knowledge
Completion method		
Telephone	Responses instantaneous High completion rate Clarify and validate data on collection No travel required – target wide geographical area	Participant needs to allocate a set period of time Confidential, quiet space required for the participant Potential for transcription errors Labour intensive May be costly if to a mobile number
Group administered	Quick turnaround – questionnaires distributed, completed and returned in one process High response rates Potential for researcher influence on the group	Small numbers – small-scale research Labour intensive Costs – associated with travel
Face to face individual	Potential to gain more depth of information Immediate validation of data Potential for researcher influence on the individual	Labour intensive Expensive – researcher time and travel costs Smaller scale research
Self-completion	No direct researcher influences Cost-effective Large-scale research	Response rates may be low Researcher not immediately available to clarify questions Limited ability to expand on the information provided

6.4.2 Web-based Electronic/Self-Completion Survey

For the advantages outlined in Table 22 a web-based electronic self-completion survey was deemed most suitable. The online survey software, Bristol Online Survey (BOS) was utilised as it complies with all UK data protection laws and is supported by Coventry University. The system is designed to support academic research allowing direct export of data into Microsoft Excel and Statistical Package for the Social Sciences (SPSS), thus saving the researcher time on data entry. Utilising an online system has several other advantages in that there is a reliable system of tracking responses. Comparative to traditional paper surveys, an electronic survey offers the ability to test and adapt the survey, provide direct links to an e-mail address, and allows document uploading. Skip sequencing is considered beneficial as it reduces participant burden; questions are tailored to meet response patterns. The participant information sheet (Appendix 6 +7) detailed the contact details for the researcher should participants require any assistance or encounter problems with the survey.

6.5 Survey Construction

The surveys were constructed with sections presented in a logical order linking themes of reporting and arbitration practices. Fink (2005) describes that a logical flow in the survey reduces the onus on participants and increases the chance of the questionnaires being completed.

To enable accurate mapping of service configuration, it was first necessary to define arbitration and consensus as the systematic scoping review highlighted the terms were used interchangeably. For this study, arbitration was classified as either a

solitary third reader who makes the final decision on their own, or via a consensus (defined as a group of 2 or more individuals) decision-making process. The surveys were divided into six main themes. Initial workforce data were requested regarding the professional background and number of reporters. Screen reading experience and professional status of staff undertaking arbitration or leading consensus meetings were also sought. Secondly, the detail of reporting practices for prevalent and incident screens was requested. The third section pertained to strategies for resolving discordant cases. In particular, why these systems were implemented and what evidence and guidance were used to endorse them. Section four related to the amount of time and scheduling afforded to third reader arbitration or consensus review. Decision-making strategies and group dynamics within consensus teams was the focus of section five. The final section of the survey was associated with implementation and current status on Radiographers' ability to comply with the recommendations within the PHE guidance.

There was a requirement to balance the survey so that the burden to participants was minimised while aiming to obtain quality data. Many factors have been associated with burden: questionnaire length, layout, format, frequency of sampling, using financial incentives and the mental onus required to complete the survey (Rolstad, Adler, and Rydén 2011, Draper et al. 2009). A strong emphasis has been on questionnaire length, with lengthy questionnaires alluded to represent barriers to completion in clinical practice with fatigue impeding the accuracy of the information provided (Subar et al. 2001 and Mark et al. 2008). Subsequently, the impetus to develop shorter questionnaires is based on this rationale. However, a review and

meta-analysis undertaken by Rolstad et al. (2011), although concerned with patient completion of surveys concluded that there is only weak evidence demonstrating a correlation between questionnaire length and response burden. The quality of the questionnaire, rather than the length, was considered the primary influence. This view is supported by Draper et al. (2009), who stated that shorter questionnaires and offering a prize draw monetary incentive did not influence responses in a postal survey of GPs. In this study, the length of the questionnaire was determined by the minimum data required to ascertain the different practices.

Specific elements of the questionnaire were mandatory, and questions were formatted to allow respondents to select from given fixed options or provide free-text comments when 'other' was selected to define this specifically (Hagell et al. 2010). Closed-ended questions facilitate the prompt accumulation of data, but as the choice of answers is determined by the researcher, the richness of data is significantly reduced (Greenhalgh 2014). Optional open elements were included throughout the survey to allow input of free-text comments to apportion scope to capture any particular opinions participants wished to express. This provided a qualitative element and richness to the survey results. A summary of the advantages and disadvantages of open and closed-ended questions are detailed in Table 23 (Boynton and Greenhalgh (2014).

Table 23 Advantages and Disadvantages of Open and Closed Survey Questions (Taken from Boynton and Greenhalgh 2004).

	Pros	Cons
Closed-ended	Appear quick and easy to complete, which may encourage participation.	Reliant on the participants understanding the question and instructions. Assumed understanding of preference or rating scales.
	Participants are not required to construct an answer	Potential for participants to randomly select an option or guess if they are unsure.
	Socially undesirable options can be included	Prone to error – the wrong box may accidentally be selected.
	Responses are generally distinct and complete.	No option for participants to elaborate on their responses or provide alternative views.
	Less effort required for standardisation, coding and analysis.	
	Suitable for self-completion or with researcher assistance.	
Open-ended	Facilitates participant originality and free articulation	Longer completion time which may deter people from participating
	Acquires responses, opinions and ideas that researchers may not have considered.	Analysis is time-consuming, requiring interpretation and coding.
	Flexibility for participants to provide as much or as little information as they desire.	Dependent on participants willingness to be expressive

Multi-item responses were also included, but the options were not ranked in a specific order. Variability exists within studies reporting that options presented first or last are more likely to be selected; with some studies reporting no influence on order at all (Krosnick and Presser 2009).

Respondents were also asked to complete their response regarding how strongly they agreed or disagreed with statements about consensus team dynamics and Radiographer arbitration. Song et al. (2015) provided a validated survey instrument designed to measure team dynamics. From the 31-items, groups were selected that were considered to reflect a breast consensus team, while maintaining the original survey constructs when feasible, for example, conditions for team effectiveness,

supportive processes, acting and feeling like a team and perceived team effectiveness. Items were excluded if they were not applicable in the context of this study (for example, changes in patient status or timely reporting of care plans). The administered survey included 15 items measuring the four factors to assess the dynamics of consensus teams, in particular similarities and differences between units relative to views of group diversity, integration, respect, accountability, and effectiveness. A five-point Likert response scale, ranging from strongly agree to strongly disagree was used. Likert Scale questions are a conventional method of collecting data, which provides a means of drawing conclusions, results and graphs from the responses.

Abbott et al. (1998) maintain that survey research can be prone to reactivity; participants complete the survey repetitively selecting the mediocre answer. Therefore, in this survey, questions requiring a response on a Likert-type scale were phrased to incorporate both negative and positive worded questions; although this method is not commonly recognised (Jones, Baxter, and Khanduja 2013). There are some studies (Diefenbach, Weinstein, and O'Reilly 1993, Russell and Bobko 1992) that advocate larger scales (7 and 9-point) should be used to increase reliability. There is also evidence suggesting that the neutral mid-option should be excluded (4-point scale) (Garland 1991). A Likert scale merely provides a rank order and does not provide a measurement of how much the responses differ (i.e. it cannot be presumed that the difference between adjacent levels is equal). However, it does provide a generalised picture of a particular topic.

6.6 Validity and Reliability

6.6.1 Validity

The validity of a questionnaire denotes its ability to measure what it claims to measure (Rebar and Gersch 2014). Jones and Rattray (2007) endorse a literature review and involvement of potential participants to substantiate face and content validity. Face validity alone is considered an inadequate measure (Svedbo Engström et al. 2018). To reduce potential researcher bias (Maltby et al. 2010) and improve the content validity, the questionnaire content was drawn from the literature, with input from a Consultant Breast Radiologist, Assistant Professor of Screening, Research Fellow, and the former Director of the NHS Cancer Screening Programmes. Constructive feedback on survey design, content, ambiguity, bias and constructs were sought, following which refinement was undertaken. One reviewer was asked to read the questions and think aloud to ascertain relevance and comprehension, as recommended in the literature (Dietrich and Ehrlenspiel 2010).

6.6.2 Reliability: Piloting the Survey

Boynton and Greenhalgh (2004) state that inapt instruments will produce poor quality data, ambiguous conclusions and result in vague recommendations. Edwards (2010) stipulated that questionnaires should be piloted; specifically gaining preliminary information on how the survey works in a realistic clinical setting, and thus adding reliability to the findings. Brace (2008) advocates pilot-testing by reviewers who represent the study's sample. The revised questionnaires were therefore pre-tested online (BOS) for clarity and comprehensive with two clinical experts of different professional roles (Director and Consultant Radiographer) from varying geographical breast screening units. The pilot study aimed to ensure that the

filtering and subsidiary questions were presented logically, response categories were compatible with the participants' experience, and sufficient space was allocated for free-text comments, thus aiming to maximise completion and value of the responses. Piloting also offered the opportunity to review the data collected, which would reveal poorly defined questions or concepts and estimate the time required for completion. One further revision was made to clarify which professional role had responsibility for the sign-off report on NBSS. In this study, the questionnaire was not re-tested as the pilot responses corroborated only minimal changes were required.

Questionnaires are considered reliable if they produce consistent results when completed by a repeat sample at a different time (Fink 2005). Variances in results are then considered trustworthy differences from the participants, rather than discrepancies in how the questions are understood or interpreted. It was not considered realistic to repeat the survey at a different time, as the purpose was to obtain a current cross-sectional response of services, and there was a conscious effort of clinician's time.

6.6.3 Response Rates

It is acknowledged that a survey does not provide a precise measurement (Salant and Dillman 1994); instead, it provides an estimate of the population in the study. Questionnaire response rates are variable, and there is no ruling on what represents an adequate response rate. Non-response rates fall into two categories; refusal to complete, or non-contact. Both elements may induce bias in the survey findings. In this study, attempts to reduce bias were achieved by distributing to all 80 Directors

within England. Also, a variety of strategies were used to promote awareness of the survey in an attempt to increase response rates. These included advertisements in Synergy news, a radiography journal automatically distributed to all Radiographers registered with their professional body (The Society and College of Radiographers SCoR) and via the Royal College of Radiologists (RCR) mailing list. A social media (SCoR Facebook) site and Twitter were also used, providing a direct link to the BOS internet Uniform Resource Locator (URL). An online Glasscubes collaboration platform was an additional source to advertise the survey to members of the Consultant Radiographer group. This approach resulted in overall responses from 61% (49) of units, deemed a moderate return (Burns et al. 2008, Burkell 2003).

Higher response rates provide external validity (Burns et al. 2008). It is acknowledged that non-response is a potential source of bias if it is deemed the non-responders are significantly different from the responders. Potentially this can dispute the robustness of the results (Atif et al. 2012) with over-representation of findings from the responders and under-representation of findings from those who do not participate. The surveys were primarily open for six weeks; a reminder e-mail containing the survey link was resent from PHE to all the breast screening units two weeks before closure, which did prompt some Directors to complete.

6.7 Trustworthiness

To increase the comprehensive and trustworthiness of a study Maltby et al. (2010) propose methodological triangulation combining quantitative and qualitative methods to obtain information about the research problem. This process may result in convergent results producing the same conclusions, which increase validity via

verification. Complementary results may emphasise different aspects of the phenomena or new phenomenon but enhance the individual results. Alternatively, it produces divergent results which may initiate new explanations (Heale and Forbes 2013).

In this study, quantitative data (national surveys, KC62 performance data) and qualitative data (telephone interviews) were utilised to provide a rich account of reporting and arbitration practices. The inclusion of verbatim quotes from free-text comments in the surveys provided depth and trustworthiness in conveying the survey findings (Moule et al. 2014). At the outset of the study, the researcher had limited experience with quantitative data collection and analysis. This was addressed by internal training courses on questionnaire development, BOS training and SPSS workshops.

6.8 Data and Statistical Analysis

All survey data were exported into SPSS® version 24 and Microsoft Excel. Descriptive statistics were used to analyse quantitative responses, including frequencies, percentages and cross-tabulations. Also, the surveys were imported into NVivo 11 (QSR International) to allow a qualitative review of the free-text comments.

6.9 Survey Results

6.9.1 Combined Survey Response Rates

33 of the 80 surveys distributed to Directors of breast screening units were completed, providing an overall response rate of 41%. As there is an unknown number of film readers (Radiologists, Breast Clinicians, Radiographers), response

rates could not be calculated for the film reader survey. Combining the results from both surveys' provided information from 49/80 units (61%). Denscombe (2014) recommends that response rates be assessed against comparable studies. A 2016 survey (Rajan and Sharma) assessing breast screening prevalent recall rates had a response rate of 49% (39/80 units).

Graph 1 demonstrates the respondent professional groups, indicating that they were a representative sample of the reporting personnel in the NHSBSP. If there was a disparity in responses within the same unit, the Director response was taken to be the instruction/process in place. Otherwise, all responses have been evaluated. Discordant views are discussed throughout the relevant sections.

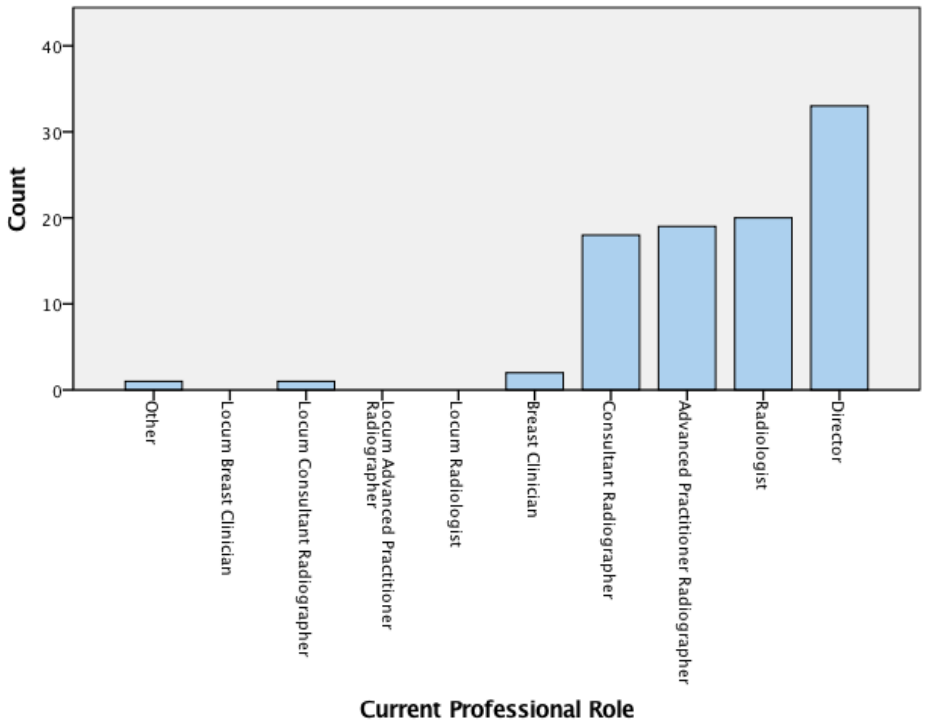


Figure 7 Graph 1. Professional Role of Survey Respondents

6.9.2 Geographic Location of Breast Screening Units

Table 24 shows the geographic location of the respondent breast screening units and the percentage of units replying. Response rates include both Director and Breast Screening Reporters surveys. The table indicates that regional response rates range from 33% to 100%.

Table 24 Number of Units Responding by Region.

Region	Units responded to survey	Total units in the region	% response rates
East Midlands	4	9	44%
East of England	6	11	55%
London	2	6	33%
North East, Yorkshire & the Humber	5	12	42%
North West	6	11	55%
South East	10	14	71%
South West	9	9	100%
West Midlands	7	8	88%
Total	49	80	61%

6.9.3 Workforce

As discussed in Chapter 1, there is a chronic shortage and a diminishing number of breast Radiologists in England. Data from this survey (Table 25) supports the RCR workforce survey (2020) highlighting that 48.5% (n=16) of the units that responded are currently operating with only one or two Radiologists (median= 3).

Table 25 The number of Radiologist's per unit (frequency) undertaking breast screening work.

No. of Radiologist	Frequency	Percent	Cumulative Percent
1	4	12.1	12.1
2	12	36.4	48.5
3	3	9.1	57.6
4	4	12.1	69.7
5	3	9.1	78.8
6	2	6.1	84.8
9	1	3.0	87.9
10	2	6.1	93.9
12	1	3.0	97.0
17	1	3.0	100.0
Total	33	100.0	

This survey did not seek to address whether the Radiologists were semi-retired or imminently due to retire, but it does demonstrate the vulnerability of some services in which there is limited provision during periods of annual leave or long-term sickness. Also, seven units were operating with one or two Locum Radiologists which may also result in less stability. The results (Graph 2) also emphasise a reduction of professionals (Radiologists, Radiographers and Breast Clinicians) with considerable (20+ years') experience in breast screening and more staff with less than ten year's practice. This rationale stimulated the researcher to investigate the potential for centralising arbitration practice, which is explored via the subsequent telephone interviews (Chapter 8).

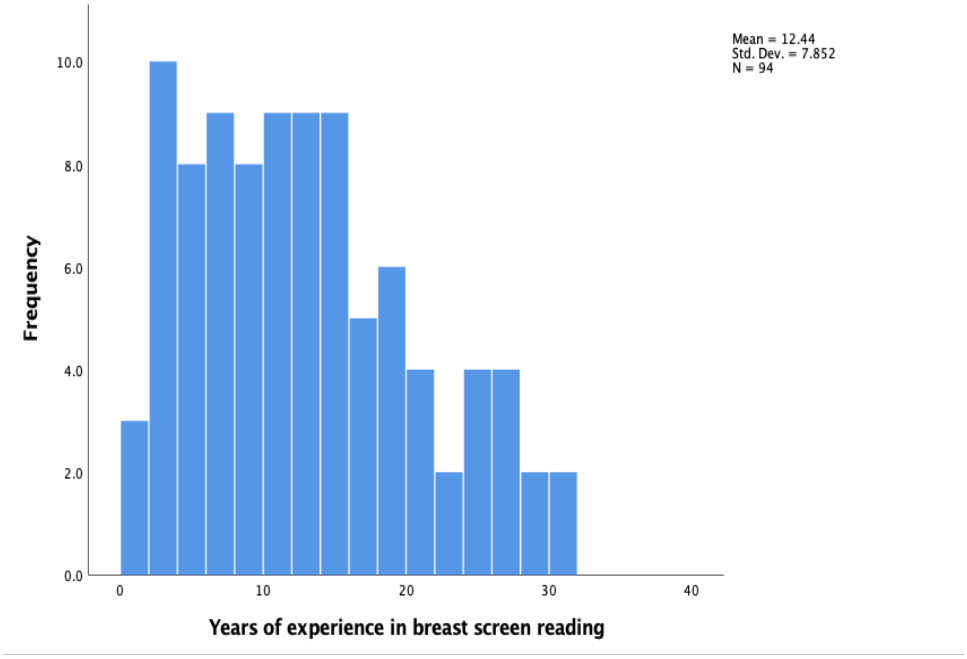


Figure 8 Graph 2. Demonstrating the Respondent's Years of Breast Screen Reading Experience.

Consultant Radiographer roles are often developed in response to several drivers, but mainly to address gaps in service delivery. All 33 Director responses confirmed that units had Advanced Practitioners, with numbers ranging from 1-6 (median=4). However, cross-tabulation of the data (Table 26) demonstrated that in those units with only 1 or 2 Radiologists (n=16) 10 had no Consultant Radiographers. These results indicate there is the potential to consider the cost-effectiveness of skill mix further in these units to support the radiology staff. Only 11 units (33.3%) had Breast Clinicians undertaking film reading, commonly employing 1 (n=7), but up to 4 in one unit.

Table 26 The Consultant Staff Structure of Units.

Count		Radiologist - Number										Total
		1	2	3	4	5	6	9	10	12	17	
Consultant Radiographer - Number	0	2	8	1	3	2	1	0	2	0	1	20
	1	2	2	2	1	1	0	0	0	0	0	8
	2	0	0	0	0	0	0	1	0	0	0	1
	3	0	2	0	0	0	0	0	0	0	0	2
	4	0	0	0	0	0	1	0	0	1	0	2
Total		4	12	3	4	3	2	1	2	1	1	33

6.9.4 Reporting Practices

When breast screening commenced, some units opted to use more than one reader from the outset. Units self-selected reading protocols (double reading with recall if one reader suggests, double reading by consensus opinion, and double reading with arbitration by a third reader) based on local operational restrictions. The reporting options currently available on NBSS are displayed in Table 27.

Table 27 Breast Screening Reading Types Currently on NBSS.

Automatic recall when unanimous (percentage)
Double reading: with automatic recall when unanimous (automatically choose opinion if unanimous)
Automatically choose abnormal (percentage)
Double reading: automatically choose abnormal (automatically choose most pessimistic)
Arbitrate when abnormal (percentage)
Double reading: with arbitration when abnormal (arbitrate unless all readers agree normal)
No automatic arbitration (percentage)
Double reading: with no automatic arbitration (direct entry but no automatic arbitration)

A seminal NHSBSP study (Blanks, Wallis, and Moss 1998) had demonstrated that double reading, and particularly double reading with arbitration, demonstrated a significant increase in the detection of small invasive cancers compared to single reading. Hence, although the NHSBSP is currently organised to operate on double reading with arbitration, an interview with the former Director of the NHS Cancer Screening Programmes highlighted that this practice had evolved naturally. Consequently, this can explain the variations in practice observed in this study.

"When the screening programme was set up originally, and I'm going back 30 years now it was set up for single reading. So, the answer is it was never set up, and people just started doing double reading however; however, they did. There was never a policy to switch the program over to double reading, and you do it like this. It was just something that grew, so people started doing it all different ways and erm that's why it is like it is now" (Former Director NHS Cancer Screening Programmes).

6.9.4.1 Blinded vs Non-Blinded Reading

To establish current reporting systems, several questions were asked about reading restrictions and practices. Responses demonstrated that the majority of units (63%, n=31) report non-blinded, i.e. the first reader's decision is visible on the NBSS reporting screen. In 29% (n=14) of units, the second reader cannot see the first reader's decision on the computer software, but this is evident from the assessment paperwork, with a minority (8%, n=4) reporting blinded double reading (Graph 3).

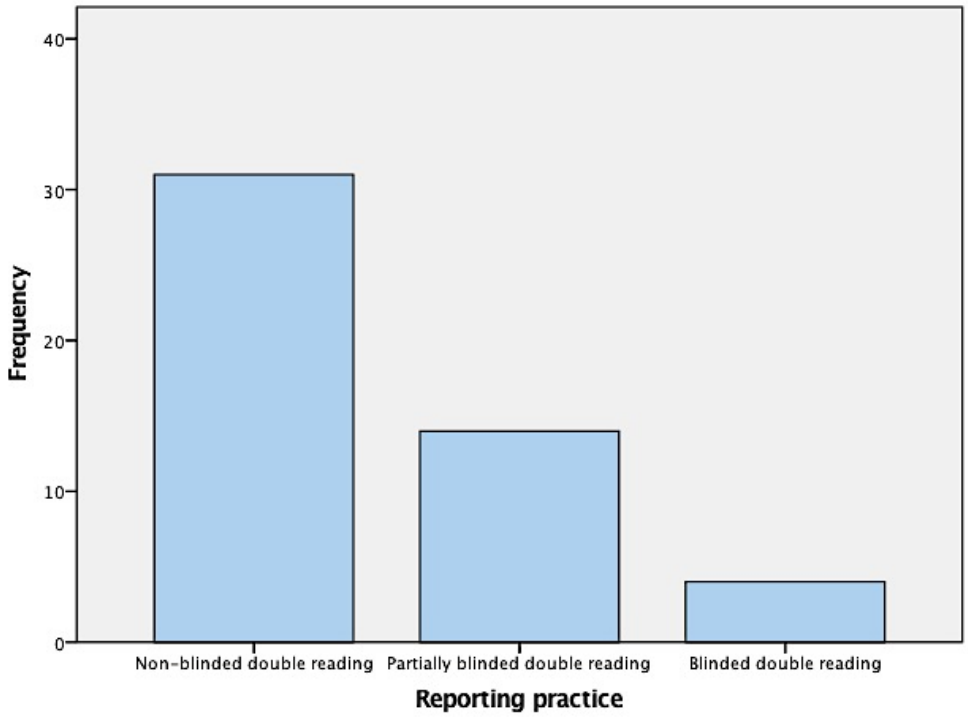


Figure 9 Graph 3. Demonstrating the Varying Reporting Practices of the Respondent Breast Units.

Analysis of free-text comments within the survey emphasised that complete blind reading is difficult to achieve with the current technology (NBSS system) requiring a paper recall system (Appendix 8 Table 28/1A).

"The screen reading workflows remain heavily dependent on paper that is difficult to both maintain and keep hidden" Director 9

However, participation in some clinical trials may require blind reading necessitating a change in local right result procedures to facilitate this.

"Would be changing to double-blinded reading when taking part in PROSPECTS" Director 30

Several individuals stated that they tried not to look at the first reader's decision as they preferred to make an independent judgement on a case and acknowledged that being aware of another reader's decision potentially influence's their decision-making.

"Our current reading practice is very biased towards the first reader's opinion. I would prefer a much more objective approach for the second reader" Film reader 1- Consultant Radiographer

For units only reviewing discordant cases the potential to bias the second reader is significant as this ultimately affects the number of cases automatically recalled to assessment. However, some survey responses indicated a benefit to non-blinded reading, describing that individuals have the opportunity to review a case and potentially change their decision or reinforce their judgement.

"It gives us time to reconsider and may cut down on discordant if we agree to recall or strengthen our opinion that there is nothing to recall and the case should go to consensus" Film reader 20 - Radiologist

Currently, the NHSBSP requires film readers to report >5000 films per year (4000 screening mammograms) including 1500 first reads. 1000 cases are a necessity for sufficient data analysis, with first reads providing true data for measuring competence. If fully blinded reading were implemented as routine practice, this would provide twice as much accurate blinded data to monitor readers and potentially change the current standards, which may be difficult for some to achieve.

"It can be hard as Radiologists to get sufficient FIRST reads…" Director 3

It is noteworthy that awareness of the staff who have reported the cases also has the potential to influence the decision made by the third reader arbitrator or group consensus.

"Stay impartial, don't let film reader names /positions affect our decisions" Film reader 41 - Radiologist

Therefore, the issue of non-blinded reading, the impact on the subsequent reader's decision-making and the limitations of the current reporting software (NBSS) were explored further in the qualitative interviews (Chapter 8).

6.9.4.2 Reporting Restrictions

It is recognised that certain combinations of readers will generate more recalls and/or arbitration cases. The NHSBSP Quality Assurance (QA) guidelines (Public Health England and PHE 2011) specify that:

"Inexperienced readers should be paired with experienced readers and, ideally, readers with high recall rates should be paired with readers who have below-average recall rates and low cancer miss rates"

In just over half of the units (51%) reporters were restricted from reading together and this was predominantly based on professional role (n=13), the experience of the reader (n=7) with minimal responses (n=5) based on individual performance data (Film Reader QA (FRQA) reports). Interestingly, one respondent stated that the

reporting restriction was dependent upon whether the majority of cases in the reporting batch were prevalent screens.

"For a clinic with over 50% prevalent screens, one of the readers must be a Radiologist with 3years experience" Film reader 54 - Radiologist

Professional role restriction was principally based on the requirement for one reader to be a Radiologist or Breast Clinician or limiting Advanced Practitioners to first reading, as demonstrated in Table 28/1B (Appendix 8).

"At least one must be a Radiologist or Breast Clinician" Director 8

While in two of the respondent units, Advanced Practitioners could not report together, they could read against a Consultant Radiographer.

"*Advanced Practitioners do not read together, but can read against a Consultant Radiographer or Radiologist" Director 2*

Even though double Radiographer reporting was endorsed in 2012 by the NHSBSP (Bennett et al. 2012) only just over half (53.1%) of the respondent units employ this as routine practice (Graph 4).

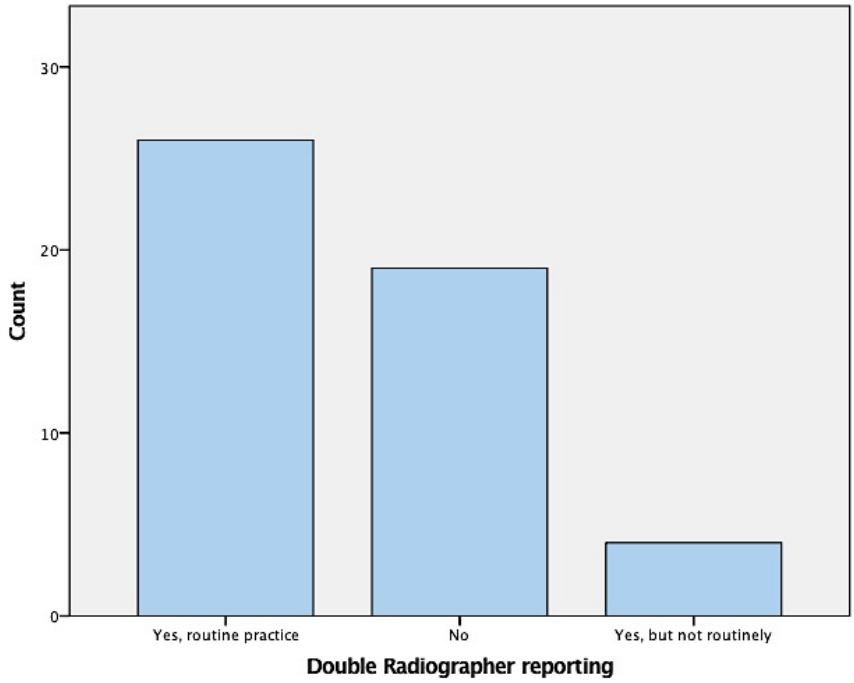


Figure 10 Graph 4. The Number of Responding Units Utilising Double Radiographer Reporting

One Director commented that this situation was historical and furthermore expressed a concern that it would be detrimental to unit performance measures (cancer detection and recall rates). However, they acknowledged they did not have data to substantiate this.

"I wouldn't necessarily say that they are considered equal to Radiologists in the sense that unit policy dictates that at least one reader is a Radiologist. This is historic and I presume is due to tradition, and the fear in the past that this might lead to an increase in recall rates and a decrease in cancer detection. I am not aware of any evidence to support this view though" Director 1.

Interestingly, one Director emphasised that although she supported double Radiographer reporting, it was the Radiographers who were reluctant to adopt this role extension.

"We would be happy to use double Radiographer reporting routinely, but the Radiographers themselves are not completely happy with this. Both Advanced Practitioner Radiographers happy to report with Consultant Radiographer but normally not with each other" Director 23

Unusually, although one unit did not utilise double Radiographer reporting, a Radiographer could undertake the third reader arbitration.

"Double Radiographer reporting (first and second reader Radiographer) not performed, but third reader can be a Radiographer if at least one Radiologist has read images as first or second reader" Film reader 25 - Advanced Practitioner

Although some units stated reporter restrictions were based on experience, the free-text comments (Appendix 8 Table 28/1B) highlighted that there was considerable variance in what units classified as an 'experienced reader'.

"Inexperienced film readers either Radiologists or Radiographers are paired with senior (>3 years' experience) Radiologist" Film reader 54 – Radiologist

"Policy going through for Radiographers of 5 years' experience to read against each other when needed only" Director 6

Only a small number of respondent units reported actively managing pairing based on individual reader performance.

"New readers do not read together for at least a year and longer if their FRQA is not within two standard deviations of the mean. Also, any outlying FRQA film readers do not read against readers with the same type of outlying reader practice" Director 16

Therefore, it appears that the pairing of readers based on individual performance measures is currently difficult to achieve and impractical in the current climate of staffing shortages and split-site working.

"We do not have the luxury of being able to pair certain readers with others!" Film reader 17 - Radiologist

Although reporting restrictions were not currently evident in one unit, the imminent retirement/semi-retirement of three Radiologists within a short period may necessitate a change as limited personnel fulfil the departmental criterion for third reader arbitration.

"The arbitration cases are done on a daily basis provided someone suitable is available. If no one is available on any given day, they wait until someone is. This may alter our practice in the future and may push us to having readers that cannot read with one another as there will be no suitable person to arbitrate" Film reader 1 -Consultant Radiographer

In the survey respondents generally categorised experience in years (ranging from 2-5), but this is debatable, and hence the classification of what constitutes an experienced reader was also explored further in the qualitative interviews (Chapter 8).

6.9.5 Current Arbitration/Consensus Practice

Information was sought for both prevalent and incident screens to determine which cases are arbitrated by a third person or reviewed at consensus meetings and the strategies used to resolve the discordant cases**.** The results (Graph 5) demonstrated national variance in whether just discordant or both concordant and discordant cases are reviewed. For prevalent screening, the majority of units (55%) review discordant only cases, but a significant number (37%) are reviewing concordant cases. The units selecting 'other' explained that technical recall cases formed part of the workload

"We have consensus meetings which include discordant recalls and technical recalls (even if both readers agreed in the technical recall)" Director 25

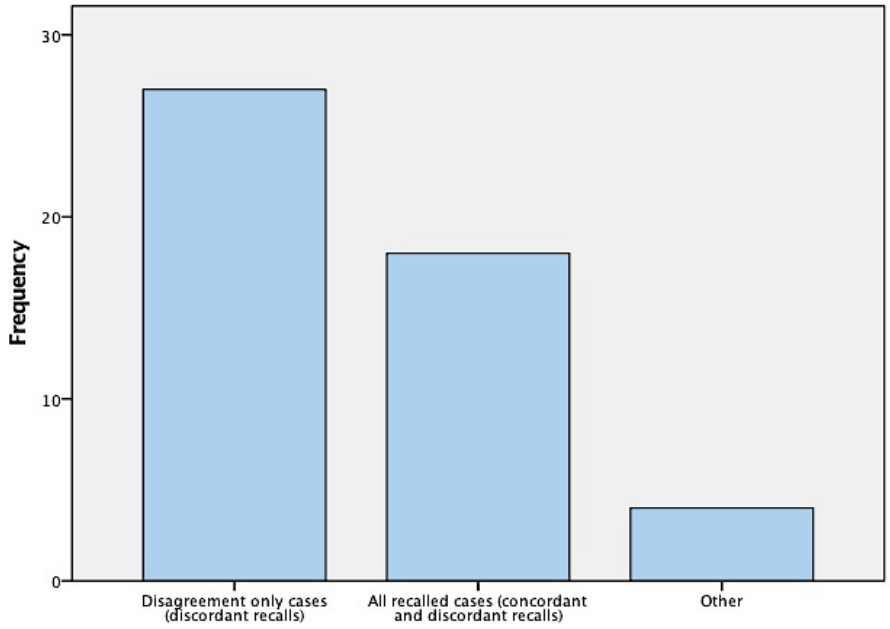


Figure 11 Graph 5. Prevalent Cases Arbitrated or Reviewed at Consensus Meetings.

For incident screening, fewer units arbitrated (third reader or group consensus) the concordant cases demonstrating units have different policies for prevalent and incident screens (Graph 6).

"We arbitrate disagreement cases only. Then, all prevalent cases that would have been recalled (all, i.e. concordant recalls and arbitrated in recalls) are reviewed in a consensus meeting to ensure prevalent recall really justified" Director 5

"Incident recall within NHSBSP standard therefore single reader arbitration. Prevalent recall as a unit, too high, therefore all recalls subject to a consensus meeting to reduce overcall" Director 27

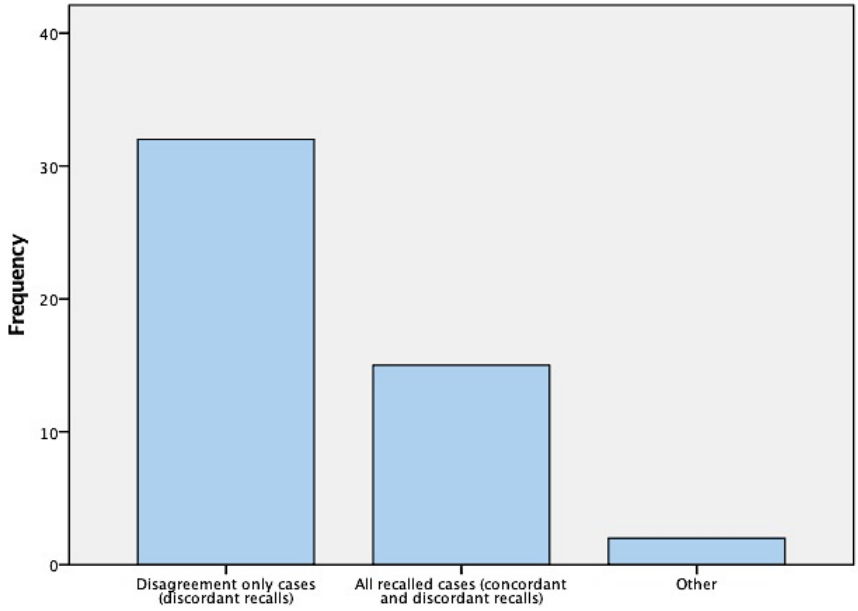


Figure 12 Graph 6. Incident Cases Arbitrated or Reviewed at Consensus Meetings.

One respondent replied that if staffing levels were not an issue, they would:

"Ideally treat prevalent and incident cases differently - our preference would be to arbitrate ALL cases recalled (whether concordant or discordant) in prevalent round" Director 24

The responses indicate that strategies have primarily been adopted in an attempt to reduce prevalent recall rates.

"We are not complying with the prevalent recall target and we now arbitrate all prevalent recalls. This is a new measure. The unit has historically struggled to achieve this target" Film reader 58 – Radiologist

6.9.5.1 Strategies to Resolve Discordant Cases

A combination of strategies is used in different units to resolve discordant cases. The graphs below (7+8) demonstrate that a consensus group which may include one or both of the original reporters is predominantly used to resolve both discordant prevalent and incident cases, followed by a single third-person arbitrator. One responding unit reported automatic recall if one reader specifies.

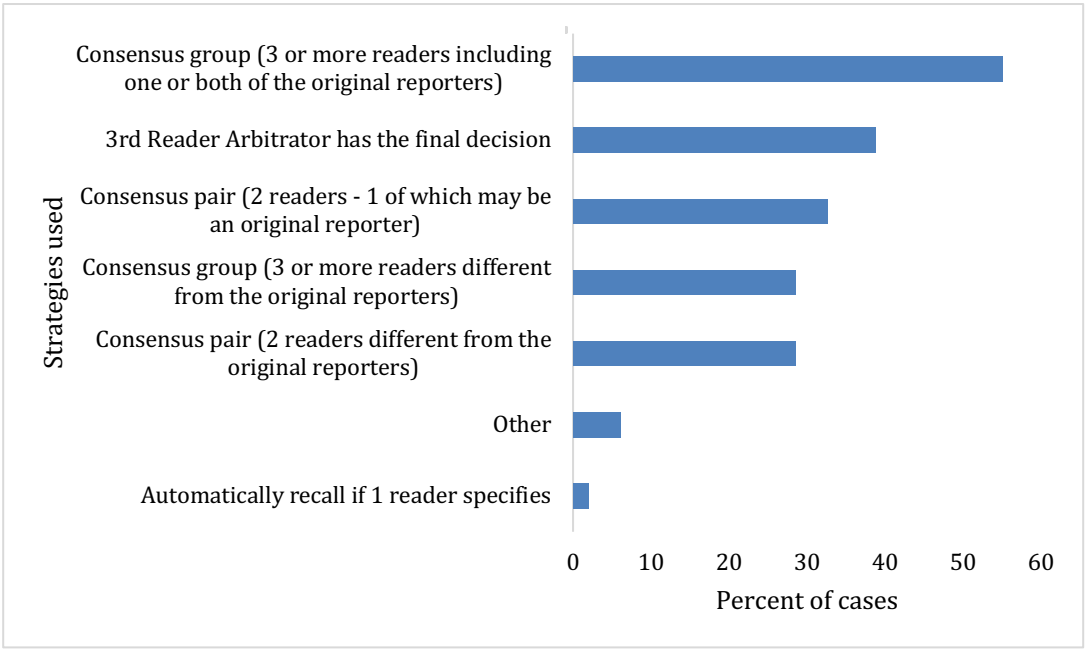


Figure 13 Graph 7. Strategies Used to Resolve Discordant Prevalent Cases

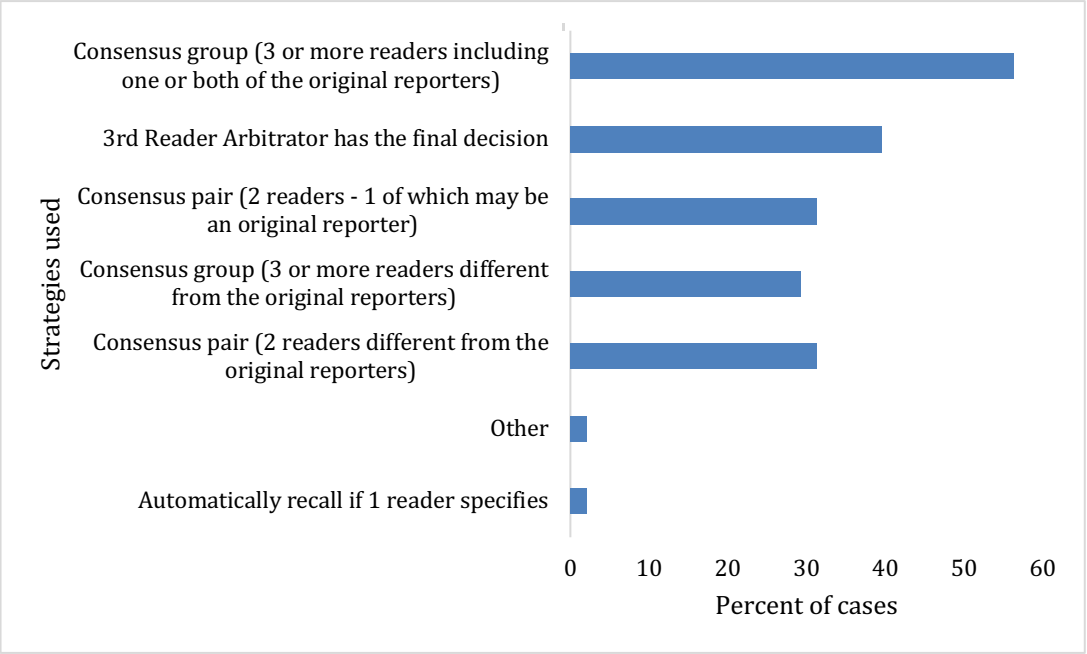


Figure 14 Graph 8. Strategies Used to Resolve Discordant Incident Cases.

The survey comments (Appendix 8 Table 28/1C) highlighted that the predominant factor determining the process was staffing resources.

"We need the flexibility of third reader arbitration in case of annual/study leave etc. where consensus is difficult" Director 24

"We would aim for consensus 3 or more readers different from the original 2 at times due to staffing we accept the other options, even 2 readers one of whom may read originally (in effect arbitration)" Film reader 61 – Consultant Radiographer

The units selecting 'other' describe a more complicated stratagem consisting of a 2-stage process involving both a third reader and group consensus.

"A third reader arbitrates all concordant recalls. If the outcome of a case is likely to be benign, that case will be put for consensus review for discussion. If the features are frankly suspicious, the third reader will ratify the recall to assessment independently" Director 9

"Third reader arbitrator first-line, but if third reader recalls then goes to consensus meeting next" Film reader 15 - Radiologist

Two respondents depicted an option of deferring difficult cases to the next day if the two/three reporters could not achieve concordance.

"The method of passing the cases to the next day's readers gives the benefit of a fresh look at challenging cases or those where agreement can't be reached" Director 1

Time constraints and the volume of cases also dictated whether concordant cases went for a review (Appendix 8 Table 28/1C).

"If time pressured and unable to consensus then if concordant for recall it is recalled" Film reader 14 – Advanced Practitioner

However, all units confirmed that they did not employ different practice dependent upon the professionals undertaking the reporting, i.e. if two Radiographers conducted the reporting, as opposed to one (or both) reporter/s being a Radiologist.

6.9.5.2 Rationale/Data to Support the Strategy Used

To determine the rationale for the strategies used to resolve discordant cases (prevalent and incident), participants were asked to provide the main reasons and to explain any supporting data. Themes for using consensus from respondents were:

- Collaborative decision-making/team approach
- Learning and educational experience
- Historical practice
- Evolution of practice
- Group opinion better than an individual
- Ensure consistent standards/openness
- Reduce recall rates – overall or prevalent
- Increase the recall rate
- Small unit -facilitates consensus with a limited number of readers

One free-text comment also asserted that consensus devolved the responsibility, alleviating the pressure of the decision from individuals.

"To reduce recalls for benign disease and reduce the number of women being recalled unnecessarily and inflicting anxiety on them. It also took the onus off one person having to make the decision alone" Film reader 42 – Consultant Radiographer

Thematic analysis of free-text comments indicated the following reasons for using a third reader arbitrator:

- Unbiased opinion
- Experienced Radiologists keep the recall rates down
- Limited staffing resources/ Workload
- Meeting NHSBSP targets to avoid breaches – arbitration performed daily
- Logistics and split-site working
- A consensus is time-consuming/resource-intensive
- Most efficient in a small unit
- Consensus did not reduce the prevalent recall rate

The above themes demonstrate that either stratagem (group or experienced third reader) has been used to try to reduce recall rates. The survey comments (Appendix 8 Table 28/1D) show that a consensus review is viewed by some as an educational and learning opportunity, providing a transparent process and the belief that a group opinion will result in a more accurate outcome than an individual opinion.

"Previously we had consensus meetings…..it was a great learning opportunity, and I feel that going from consensus to arbitration instead has been a step backwards really" Film reader 53 - Advanced Practitioner

"A group opinion is better than that of an individual" Director 17

However, divergent views were also apparent, demonstrating a conformity of reader practice and cultural dynamics.

"I think it affects reading - if you know your recall will not get through arbitration do you stop recalling similar things? I think it negates the point of double reading

in that differences in how people read is how single reader cancers get picked up. I wish we could just do a third read in a dark room" Film reader 12 - Radiologist

"In another unit, I worked for some time in the Arbitration meeting was very dominated by the Unit Director and their opinion usually prevailed" Film reader 17 - Radiologist

It was notable that consensus meetings are utilised in different ways. One respondent described that normal and benign assessment cases were also reviewed in consensus meetings; therefore, using this as a formal method of feedback. The consensus was also portrayed as a filtering mechanism with readers sending any cases they would like to be discussed, knowing that it was not necessarily going to influence the unit recall rate. However, this system of working does influence the individual readers recall statistics and subsequent performance outcomes.

"A relatively high proportion of cases are recalled at first read, and there is no real pressure on readers not to recall for consensus/ arbitration. It is not uncommon that a case recalled by two readers will be put to routine recall following consensus" Film reader 7 - Radiologist

Conversely, third reader arbitration was deemed to represent an unbiased opinion that is not influenced by more vocal team members or by professional roles.

"The reader who recalled it is often present - whoever shouts loudest gets their case back" Film reader 12 – Radiologist

"Radiographers easily swayed by the doctors!" Director 8

Several respondents believed this was the most cost-effective (time and resources) for their unit, and a solitary third read was often employed due to staffing shortages, logistics of cross- site working and job plans. One of the NHSBSP quality standards is that ≥ 95% of women who do not require further tests are sent their result within two weeks of screening and > 98% who do require further tests are offered an appointment at an assessment centre within three weeks. Consequently, the Directors were asked if their unit had failed screen to assessment or screen to routine recall as a result of cases awaiting arbitration. From the 33 Director responses, 17 (51.5%) selected this had happened on occasion within the last five years, free-text comments stated this was usually during peak holiday times or awaiting previous images. However, this may increase with the reduction in professionals to undertake the task.

"Retirements and work patterns can cause delays in arbitration taking place" Film reader 1 – Consultant Radiographer

Interestingly, only one unit had sent cases externally to another breast screening service to be arbitrated, and this was due to no Radiologist being available during an annual leave period. The availability of previous films is an essential factor for decreasing inappropriate recalls and appreciating the subtle but significant change. In units that are struggling to report on time, awaiting acquisition of images from external centres can lead to a failure of the quality standards.

"Policy on whether one waits for priors" Director 30

Only 23/33 of the Directors completed the question which asked for any data used to support the arbitration system implemented in their unit. The vast majority (61%) responded that no data was used to support the system in place. Of the nine completed responses, statements related to review of unit/film reader statistics (n= 6), audit advised by QA (n=1) and local review of Advanced Practitioner/Consultant Radiographer film reading and arbitration data (n=2) showing parity to Consultant Radiologists.

"I have reviewed and circulated film reading data (volume, recall rate, cancers detected) for the last decade in a format where the individual can see their data compared to their colleagues, as well as their own linear data so we can look for trends. If I see changes in a reader (e.g. drop in CDR), we discuss strategies to optimise. If I see changes in unit performance - e.g. prevalent round recall rate, we discuss as a unit, audit, present data, review examples and try to change as a team" Director 2

Four units had trialled consensus meetings either for prevalent screens (concordant and discordant) or for all recalls but reverted to their original practice as it incurred delays in the final report; the reduction in recall rates was not realised or not sustained (Table 29).

Table 29 Reasons Units Changed Strategies and Reverted to Their Original Practice.

Strategy trialled	Reasons for reverting back
3-month trial trying to consensus ALL recalls (at the request of QA Radiologist).	• Disaster - could not keep within NHSBSP targets so reverted to reviewing discordant cases only
Concordant prevalent round arbitration introduced a couple of years ago.	• Initially improved recall rates in this group. A benefit was no longer evident • Exercise considered rather wasteful of limited time resources. • Replaced the regular consensus meeting with third reader arbitration and separate educational activities. Resulted in reductions in delays, but the educational aspect has never been resourced.
For a time, had a consensus meeting for all prevalent recalls. The consensus meetings were established to reduce the number of unnecessary recalls for first-timers.	• Very time consuming, required multiple people to be available at the meetings, very problematic as split-site working. • Meetings likely included one or both of the original readers who rarely changed their opinion. Overall very few women were subsequently returned to routine recall, and so the practice was discontinued. • An unwritten rule that any person making the final decision should be able to perform an assessment workup and reach an appropriate conclusion; therefore, film readers do not currently arbitrate
Previously used consensus for all first screen recall	• Limitations of staff availability led to unacceptable delays in final reporting so reverted to third reader arbitration

6.9.6 Standard Operating Procedures (SOPs)

A crucial feature of any Quality System is working in accordance with standardised and clear Standard Operating Procedures (SOPs). Working from an agreed and formal process has the potential to reduce the risk of errors and establish consistency regardless of variance in who is undertaking the task. 30 (61.2%) units responded that they adhered to SOPS for arbitration, 9 (18.4%) units responded no, with 10 (20.4%) who did not know. However, it was apparent that there were differences in responses from the Director and film readers, and between peers within the same unit (Graph 9). This may reflect a communication issue within some centres where policy has not been disseminated.

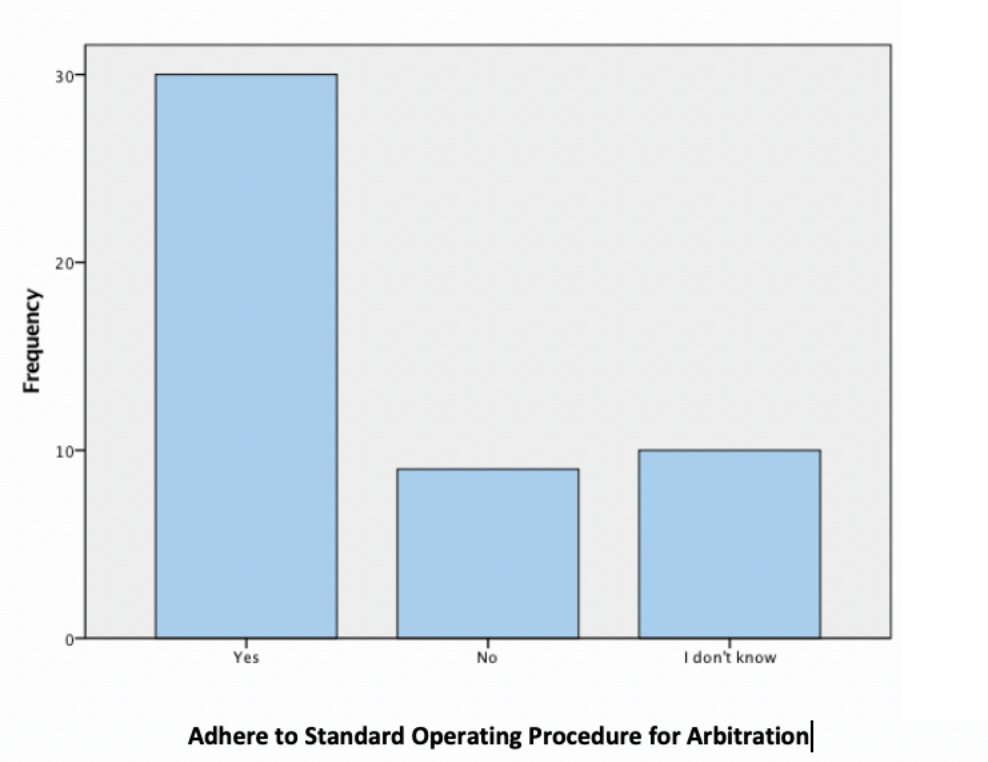


Figure 15 Graph 9. Unit's Adhering to Standard Operating Procedures for Arbitration.

6.9.7 Scheduling of Arbitration/Consensus Meetings

Several questions in the survey attempted to understand the resource implications (time, clinical resources, number of sessions and cases) involved in consensus review meetings. It was acknowledged that there would be a variation in the number of cases requiring review every week due to the prevalent/incident ratio. However, the aim was to gain an overview of whether the time/resources invested in consensus, and in particular, the review of concordant recalls, correlated with units recall rates. This is explored further in Chapter 7.

Primarily the survey assessed whether third reader arbitration or consensus review meetings were scheduled or occurred ad-hoc. In the majority of cases (63.3%), there was no scheduled time (Appendix 8 Table 28/1E).

"We do not have the luxury of scheduled protected time for arbitration/case discussion which we all find extremely frustrating and know is not ideal" Director 24

Moreover, the predominant determining factor for when it was performed was again the availability of sufficient staff with the relevant clinical skills (Appendix 8 Table 28/1E).

"When staff are available- often staff shortage, just doing our best to get it done asap!" Director 21

"When the Radiologist has some free time to do it" Film reader 53 – Advanced Practitioner

It was apparent from the survey comments in Table 28/1E (Appendix 8) that clinical workload impacts on the opportunities to review the cases with some units balancing competing demands which is proving particularly difficult.

"It is a constant pressure, but seen as highly valuable" Director 2

"Scheduled time for this was tried but failed as invariably personnel were required for clinical needs" Film reader 8 – Consultant Radiographer

In those units with dedicated time for a review of cases this was either first thing in the morning before a clinic commencing, or after Multidisciplinary Team (MDT) meetings to maximise the number of staff available.

"Consensus meetings are at the same time each week, so everyone (readers and office) know when they are" Director 13

One-unit utilising both a third reader arbitrator and group consensus, had a weekly team meeting reserved to discuss the more complex cases (integrated screening and symptomatic patients).

"Third reader arbitration may occur at any time during a working week according to the availability of the Radiologists. We have a single consensus meeting to discuss more complicated screening and symptomatic cases, cases from third read prevalent arbitration and for housekeeping outstanding incident discordant cases" Director 9

In another unit, two or three readers report all films for a particular day (usually two working days after the images were obtained). At the end of the day, these readers hold a consensus meeting. If there is no explicit agreement, then the case is passed for arbitration to the two/three readers on for the following day. Although this system guarantees a timely review of cases, it is entirely dependent upon having sufficient film readers.

Interestingly, only one respondent mentioned the time of day as a factor for performing arbitration. However, nothing in the literature review highlighted the time of day for reviewing complex cases when attentiveness may not be optimum.

"Try to avoid doing at the end of the day due to time and alertness, but this happens if necessary" Film reader 5 – Advanced Practitioner

The number of sessions to undertake third reader arbitration or group consensus varied greatly, ranging from 1 to 8. The time dedicated to reviewing cases was invariably dependent upon if it was a solitary third reader who may be undertaking arbitration within their dedicated reporting session or a group review. However, in a typical week, the time was stated to range from 15 minutes up to 6 hours, and subsequently, the number of cases reviewed was wide-ranging (4-200).

"Takes a lot of time up of all our reading staff to meet every morning" Film reader 12 - Radiologist

The reported group size was also highly variable with consensus undertaken in pairs, while another unit adopted all members of the team attending.

"All team members are at consensus unless on leave, /still in clinic" Director 13

Across the remaining programmes, group sizes ranged from 3 up to 10.

6.9.8 Consensus Practice

PHE acknowledges that arbitration may be undertaken in different ways;

"A third image reader or a small group or panel of image readers to arbitrate on these cases".

The introduction to the survey highlighted that for this study, arbitration was classified as either a single arbitrator (third reader) and a consensus was any form of group/pair review of cases. When respondents were asked if they undertook any form of consensus (group/pair) a number (10) answered no, but in free-text, comments described a group process. This emphasises a current problem in understanding the terminology of arbitration and consensus, with the two interchanged, and this was reflected in the survey results with individuals from the same units selecting differing practices. Hence, these ten respondents were not directed via the routing within BOS to the consensus sections within the survey. In total, 66 respondents of varying professional roles completed this component of the survey.

To establish a clearer picture of consensus review, various questions were posed about team membership, decision-making strategies and team dynamics. To understand the group composition of consensus (grade and /or the number of staff), respondents were asked to specify if a minimum (Quorum) membership was

required for the meeting to go ahead. Of the 36 programmes which undertook some form of consensus review, 25 required quorate and 11 did not (Graph 10).

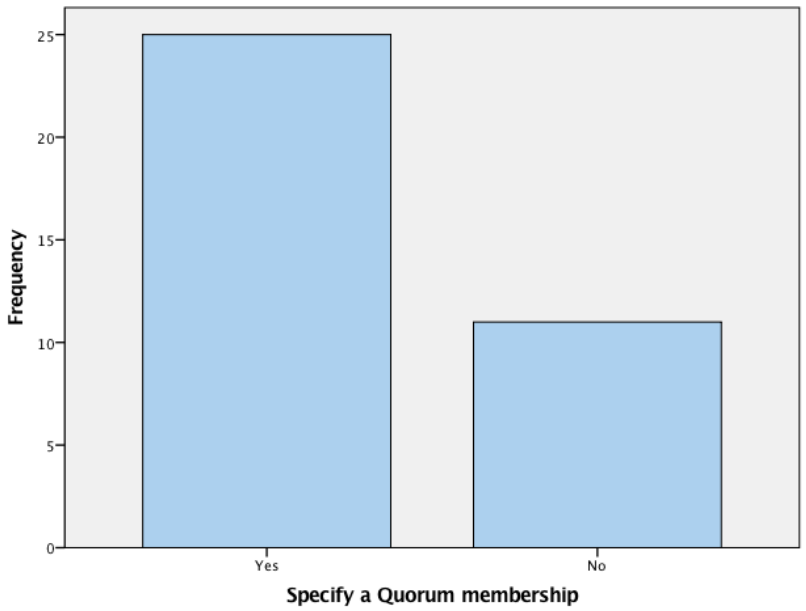


Figure 16 Graph 10. Necessity for Quorum Membership.

Quorum requirements varied (Appendix 8 Table 28/1F) ranging from merely specifying the number of staff that must be present (predominantly 2 or 3), stipulating the presence of a particular professional role, to mandating that the group must consist of staff who had not already reported the cases.

"Minimum of 2 and one of them has to be Radiologist" Director 28

"Consultant Radiologist or Consultant Radiographer must be present" Director 16

"Two people not involved in the 1st/2nd read" Film reader 46 – Advanced Practitioner

6.9.9 Decision-Making Strategy at a Consensus

As discussed in Chapter 4, group decision-making can be achieved in several ways. Consensus denotes that discussion results in the group achieving a decision. Graph 11 demonstrates that in the main (69.4%, n=25) a majority decision (equal skills assumed) is used in consensus meetings. In 13.9% (n=5) of units, the decision to recall was weighted by experience, and in 2 units (5.6%) this was undertaken if any individual specified recall.

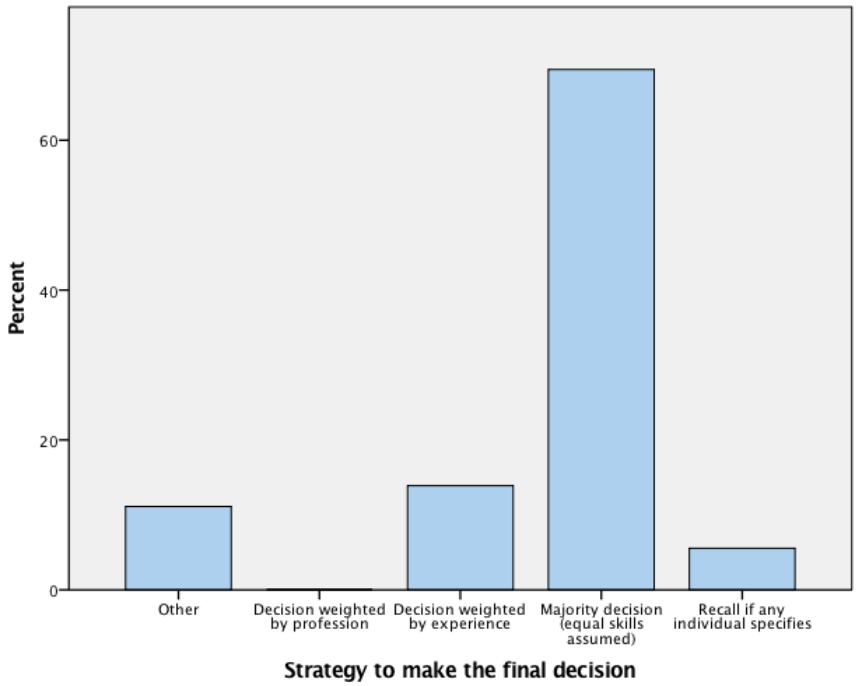


Figure 17 Graph 11. Decision-Making Strategies at a Group Consensus

Although some units reported that a majority decision was the strategy used, four responses selecting 'other' highlight that decision-making is multifaceted and may include many different options and recall would be instigated if one individual had strong views on a particular case. This likely explains why there were conflicting responses between peers in the same unit (Appendix 8 Table 28/1G).

"If one reader is really concerned and asks that the patient be recalled then despite the consensus group not wanting to recall - the patient will be recalled" Film reader 42 – Consultant Radiographer

A strength of opinion could also be conveyed to the consensus group by individuals documenting this on the assessment recall sheet (Appendix 8 Table 28/1G).

"If the film reader/Radiologist/ Breast Clinician wants the cases to come back for further views regardless of the consensus process they write on the paperwork" Film reader 3 – Advanced Practitioner

Although opinions were deemed equally weighted in principle, some respondents acknowledged that experience is an influential factor (Appendix 8 Table 28/1G).

"*Two readers are usually able to reach an opinion, regardless of pairing, but I am sure that experience carries weight in practice" Director 2*

Other comments described that it is a different decision-making process when reviewing cases in consensus compared to film reading and on reflection, individuals may alter their original decision.

"I find that I approach consensus with a different mindset to reading, and will often change my opinion at consensus after deliberation and discussion" Director 2

It is recognised that a different skill set is required when arbitrating; in particular, there is a need to be highly specific and hence reduce unnecessary recalls.

6.9.10 Team Dynamics Within Consensus Groups

As discussed in Chapter 4, a key element of consensus is the disposition of the team members and the ability to actively listen to other viewpoints. Survey comments in Table 28/1H (Appendix 8) demonstrate a positive team culture in some units with collaborative working

"We are fortunate to have a very strong team of film reading/Consultant Radiographers who are held in very high regard by their Radiologist colleagues, and I do not have any sense that their opinions are not highly valued. The Consultant Radiographers are in the unit 5 days a week, unlike Radiologists who have other roles outside the department, and hence are very good at maintaining consistency" Director 2

Conversely, some free-text comments (Appendix 8 Table 28/1H) supported evidence from the literature review (Bankier et al. 2010, Wolf et al. 2015) regarding the complexities of dynamics that exist within teams where one member is dominant, and individuals with strong influential characteristics drive decisions.

"Depends on the mix of staff in the group on the day as to how valued and respected any member's opinion is. The day I generally attend (I am part-time) the consensus is led by a Radiologist who does not generally value other opinions but if it is led by a different Radiologist or Consultant Radiographer all opinions are valued and treated with respect" Film reader 51 – Advanced Practitioner

To gain an understanding of how team dynamics affect consensus group meetings respondents were asked to rate on a 5 point Likert scale how strongly they agreed or

disagreed with 15 items (0=strongly disagree, 1=disagree, 2= Neither agree nor disagree, 3=Agree, and 4=strongly agree). There was only one negatively worded item which was reverse scored. Survey properties assessed were means, minimum, maximum, and variance.

The survey responses suggest that on average respondents perceived a neutral to a positive level of team dynamics within consensus groups (Table 30). Mean scores for individual items ranged from 1.56 to 3.52, indicating a potential for some improvement in team dynamics measured by this survey. The item with the most positive response was 'Consensus meetings provide an opportunity for educational learning from cases. The item with the least positive response was 'Membership of the consensus group changes frequently, so there is not a set team'. The conceptual model developed by Song et al. (2015) states that a stable team is one of the three enabling conditions for increased effectiveness. In a diagnostic setting, changeable team membership is recognised as a strategy when pre-planned coordination is not feasible (Bushe and Chu 2011 and Vashdi et al. 2013). Although instability of a team can be associated with a diminished sense of belonging for the individuals, (Bushe and Chu 2011 and Shumate et al. 2010) in a breast consensus group variation of team members can be beneficial as individuals have different aptitudes for detecting the differing types of mammographic abnormality (for example, architectural distortions). This lower score is, therefore not considered detrimental to the team dynamics in this study. However, the next least positive response was '*Our team has mechanisms in place to monitor consensus outcomes*' which suggests in some units a lack of audit and feedback.

Nevertheless, it is essential to note that overall responses were variable with 8 out of the 15 items demonstrating a polarised opinion. This is demonstrated by the minimum and maximum scores displayed in Table 30. One unit described a recall book, and the outcomes of the consensus group are documented, which provides a feedback mechanism for team members who are not present at the meeting.

"If a team member who recalled a case is NOT present and the consensus is to return to screen, a note is made of their initials on the margin of our "recall" book for that case so that reader can review decisions against their suggested recall"
Film reader 20 - Radiologist

Although the number of negative responses in the survey was small, they related to all items in the factor 'Process for communication and information exchange', and specific items within 'process for conflict resolution', 'acting and feeling like a team' and 'perceived effectiveness'. This implies that team dynamics are problematic in individual units, and a consensus is only constructive in specific settings.

Table 30 Team Dynamic Factors with Mean values and Standard Deviations of the Conceptual Variables

Factor Name	Item number and Text	Mean	SD	Min.	Max
Conditions for team effectiveness	Membership of the consensus group changes frequently so there isn't a set team*	1.56	1.490	0	4
	The consensus group has the right "mix" of staff—a group of people who bring different clinical perspectives and experiences to the discussion	3.29	.602	2	4
Shared Understanding	There is a real desire among team members in the consensus group to work collaboratively	3.39	.782	1	4
Process for accountability	Each group member shares accountability for consensus group decisions and outcomes	3.26	.810	1	4
Process for communication and information exchange	Consensus meetings provide an open, comfortable, safe place to discuss cases	3.32	.963	0	4
	Consensus meetings provide an opportunity for educational learning from cases	3.52	.707	0	4
	During the meeting, team members ask for and give each other constructive feedback.	2.82	1.036	0	4
	Our team has mechanisms in place to monitor consensus outcomes	2.76	1.039	0	4
Process for conflict resolution	When team members disagree, all points of view are considered before deciding on the final outcome	3.33	.810	0	4
	Within the consensus group, we are able to work through differences of opinion without damaging relationships	3.35	.774	1	4
Acting and feeling like a team	Members of the consensus team depend on each other for their special knowledge and expertise	3.30	.744	1	4
	Members of the consensus group show respect for each other's roles and expertise	3.35	.813	0	4
Perceived team effectiveness	The way the consensus group members interact improves the quality of patient care	3.26	.900	0	4
	I feel integral to the consensus group	3.38	.799	1	4
	I experience excellent teamwork with the members of the consensus group	3.30	.803	1	4

*Reverse-coded

Cronbach's alpha for all variables was 0.915 representing excellent internal consistency. Mean scores were calculated for the five multi-item factors. Computing Cronbach's alpha coefficients assessed factor reliability. The internal consistency is categorised into five categories, as demonstrated in Table 31. Four of the five multi-item factors showed acceptable to excellent reliability ranging from 0.676 to 0.920. The one factor scoring < 0.5 was 'Conditions for team effectiveness', but as previously discussed, this measures a team's stability by a fixed membership.

Table 31 Cronbach's Alpha Scores Relative to Internal Consistency (Taken from Rose Jemutai and Wambua 2016)

Cronbach's alpha	**Internal consistency**
$\alpha \geq 0.9$	**Excellent** (High-Stakes testing)
$0.7 \leq \alpha < 0.9$	**Good** (Low-Stakes testing)
$0.6 \leq \alpha < 0.7$	**Acceptable**
$0.5 \leq \alpha < 0.6$	**Poor**
$\alpha < 0.5$	**Unacceptable**

Correlations among factors were moderate, averaging 0.56. There was a strong positive relationship apparent between communication and perceived team effectiveness (r = 0.845) and acting and feeling like a team and perceived team effectiveness (r=0.822) as demonstrated in Table 32.

Table 32 Data Demonstrating the Cronbach's Alpha Score and Correlations Among the Variables for Team Dynamics.

		Factor correlations						
Factor	**α**	**1**	**2**	**3**	**4**	**5**	**6**	**7**
1. Conditions for team effectiveness	0.161	**1**						
2. Shared understanding	*	0.337**	1					
3. Process for accountability	*	0.180	0.503**	1				
4. Processes for communication and information exchange	0.676 Acceptable internal consistency	0.149	0.664*	0.531**	1			
5. Processes for conflict resolution	0.914 Excellent internal consistency	0.268*	0.613**	0.507**	0.699*	1		
6. Acting and feeling like a team	0.717 Good internal consistency	0.148	0.499**	0.541**	0.767**	0.771**	1	
7. Perceived team effectiveness	0.920 Excellent internal consistency	0.203	0.679**	0.598**	0.845**	0.783**	0.822**	1

*Cronbach's alpha not reported as the scale is based on a single item

**. Correlation is significant at the 0.01 level (2-tailed).

To determine if there were differences in responses between professional roles, a Kruskal- Wallis test was conducted. The results in Table 33 demonstrate a significant difference in response between Advanced Practitioners and Directors regarding 'Members of the consensus group showing respect for each other's roles and expertise' (H (3) 10.19, p=0.017). Advanced Practitioners scored this variable low compared to the Directors scoring it high, demonstrating a difference in how professional roles perceived they are valued.

There were also very weak differences (H (3) 7.81, p=0.05 and H (3) 7.75, p=0.052 respectively) in responses between Radiologists and Directors for the two categories of 'all points of view are considered' and 'constructive feedback given' with the Radiologists scoring this lower. Potentially, with more participants, these two categories may also demonstrate a significant difference.

Table 33 Differences in Team Dynamic Responses From all Professional Groups

Team dynamic questions	P-value – comparing responses from all professional roles
Membership of the consensus group changes frequently so there isn't a set team*	0.823
The consensus group has the right "mix" of staff—a group of people who bring different clinical perspectives and experiences to the discussion	0.085
There is a real desire among team members in the consensus group to work collaboratively	0.064
Each group member shares accountability for consensus group decisions and outcomes	0.235
Consensus meetings provide an open, comfortable, safe place to discuss cases	0.172
When team members disagree, all points of view are considered before deciding on the final outcome	0.050
During the meeting, team members ask for and give each other constructive feedback	0.052
Our team has mechanisms in place to monitor consensus outcomes	0.897
Within the consensus group, we are able to work through differences of opinion without damaging relationships	0.131
Members of the consensus team depend on each other for their special knowledge and expertise	0.146
Members of the consensus group show respect for each other's roles and expertise	**0.017**
The way the consensus group members interact improves the quality of patient care	0.252
I feel integral to the consensus group	0.070
I experience excellent teamwork with the members of the consensus group	0.186
Consensus provides an opportunity for educational learning from cases	0.601

6.10 Public Health England Arbitration Guidance

To ascertain the impact of the PHE arbitration guidance, various questions were asked to establish which professionals undertake solitary third reader arbitration/lead consensus, if this practice was implemented before the guidance, and if so, what criteria the unit Director had used to determine an individual's

suitability. Subsequently, the survey assessed if the guidance had/or will change practice in units.

6.10.1 Professionals Currently Undertaking Third Reader Arbitration

Graph 12 demonstrates that third reader arbitration (professionals make a final solitary decision) is predominantly undertaken by Radiologists (51% of cases), but interestingly slightly more Advanced Practitioners carrying out this task compared to Consultant Radiographers (14.3% vs 12.2%). Two responses in the other category specified that third reader arbitration was rarely used or as an emergency measure. Although third reader statistics can be run from the FRQA, performance measures are not reported to the same extent as first and second reads. As discussed in Chapter two, regional data has (*Symposium Mammographicum Conference 2016*) demonstrated that third person arbitration results vary widely depending upon the individual undertaking the task. Hence arbitration has the potential to significantly affect how many assessment clinics are required and how many cancers are detected. Only one free-text comment specified the requirements that a Radiologist must meet in order to perform arbitration.

"To arbitrate a Radiologist must have three years' film reading experience and consistent recall rates below minimum" Film reader 54 - Radiologist

One Breast Clinician commented that some Radiologists shun arbitration cases which places stress on colleagues. Subsequently, this limits the number of personnel undertaking the task and, as all readers have personal blind spots, this may exacerbate incorrect decisions (recall or not).

"Some Radiologists avoid it, which increases the pressure on those doing it and the problem of reader weak points being overexposed" Film reader 13 – Breast Clinician

The complexities of defining quantitative guidelines for third reader arbitration were subsequently explored in the qualitative interviews (Chapter 8)

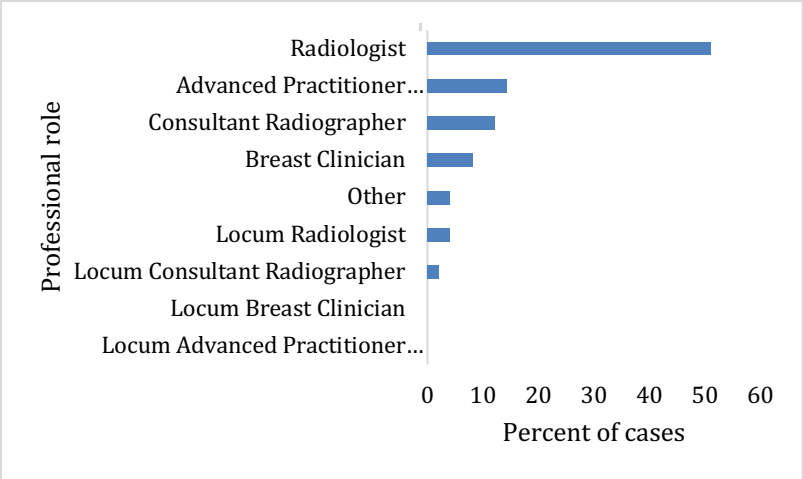


Figure 18 Graph 12. The Professional Roles Currently Undertaking Third Reader Arbitration Based on the Returned Survey's

6.10.2 Professionals Currently Coordinating/Leading Consensus Meetings

In units adopting a consensus review, the PHE arbitration guidance states that the delegated individual may be the coordinator or lead of such a group. Again, there was variation between the Director response and film reader response (n=6). Seven units selected they had no lead (Graph 13). However, free-text comments clarified it was considered a joint or group decision with parity of staff present (Appendix 8 Table 28/1I).

"We arbitrate as pairs – there is no lead, both individuals take responsibility, and both are captured on NBSS" Director 2

The one participant selecting the response 'other' described that consensus was done in pairs, one of whom is always a Radiologist.

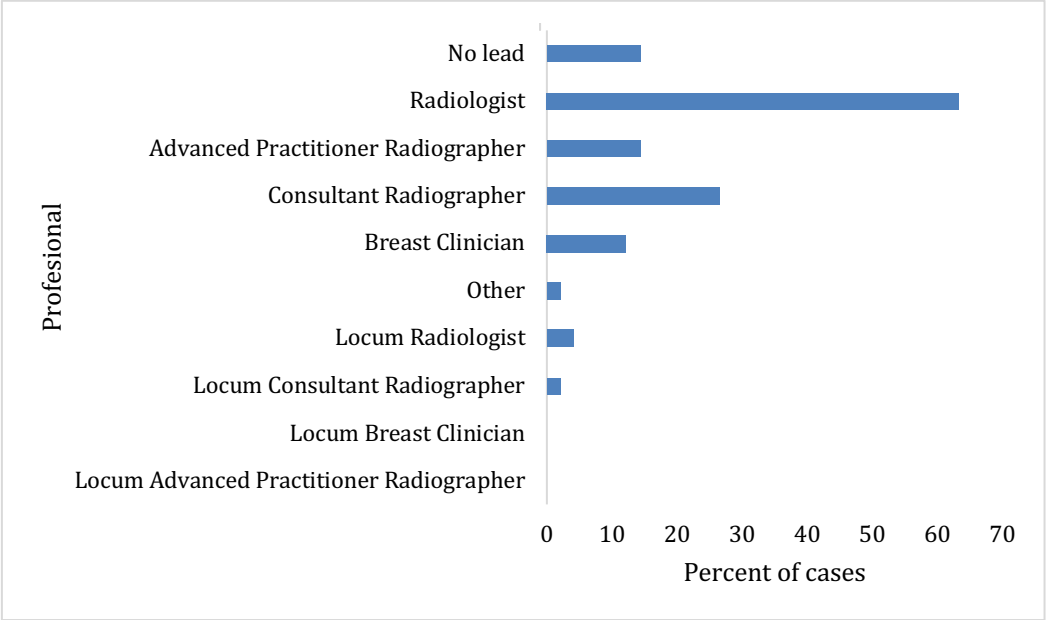


Figure 19 Graph 13. The Professional Roles Currently Coordinating/Leading Consensus Review Meetings

Interestingly, although the lead was predominantly still a Radiologist (63.3% of cases), there was a much higher proportion of Consultant Radiographers (26.5%) undertaking this role compared to third reader arbitration.

Several free-text comments in Table 28/1I (Appendix 8) highlight that there is no standardised way of recording of the consensus group decision onto the NBSS system. Some units are entering under the Radiologist's code (leading the meeting), with others entering under a group code which did not capture the personnel

present during the discussion. Interestingly, one unit captured this information via unique reader codes, which is then a valuable electronic record for audit purposes.

"This is captured on NBSS as numerical pairs of coding, e.g. 1011" Director 2

6.10.3 Timeframe for Radiographer Arbitration/Lead of Consensus

In units utilising Radiographers to perform third reader arbitration or lead/co-ordinate consensus meetings, there were limited responses as to when this policy was established. However, the survey comments in Table 28/1J (Appendix 8) indicate that some units adopted this practice several years before the guidance supported through local trust governance. One respondent commented that Radiographers were stopped from arbitrating and then recommenced when the guidance was issued.

"Since my appointment two years ago, this was underpinned by the Trust until the NHSBSP guidance was published allowing Radiographers to arbitrate" Film reader 1 – Consultant Radiographer

6.10.4 Criteria for Delegation of Third Reader Arbitration/Consensus Lead to Radiographers

To establish the inclusive aspect of delegation, it was considered necessary to explore what criteria units had used to determine an individual was suitable to undertake third read arbitration or lead/coordinate a consensus meeting. The requirements varied ranging from purely years of experience as a film reader, an expectation as part of a Consultant Radiographer role, to full compliance with all the parameters. These are detailed in Table 34.

Table 34 Criteria Used to Delegate Third Reader Arbitration/Consensus Lead.

Criteria	Example of responses
Experience	Experienced film readers Expertise, time duration as an image reader Fulfil departmental protocol with regard to years of experience reading
Unit policy as a Consultant Radiographer	It is part of my role Works autonomously within breast imaging Consultant Radiographer only Being Consultant Radiographer
Education	Master's degree required Experience and qualifications
FRQA data	NHSBSP film reading statistics Excellent proven track record of FRQA data in cancer detection & recall rates Sensitivity being over 90%
PERFORMS data	Performs performance
Number of films read per annum	Minimum 5,000 reads per year.
Audit/Interval cancer review	Participate in an Arbitrated Cancer audit with continuous assessment and feedback, and Interval Cancer reviews
Responsible assessor in assessment clinics	Undertaking assessment clinics Acted as the responsible assessor in assessment clinics
Active participation in MDT's	Actively participate in MDT meeting
Accreditation	Accredited practitioner

Interestingly, one response denoted that upon qualification as a film reader; there was no distinction from the other professional roles undertaking arbitration/consensus.

"From Day 1 of reading, Radiographers are regarded as being of equivocal status to all other readers and perform arbitration" Director 7

In those units that have recently adopted this practice, free-text comments confirmed they were utilising the PHE guidance.

"I had to meet the recommendations in the NHSBSP publication "Guidance on who can undertake arbitration (2016)" Film reader 27 – Consultant Radiographer

"The person that arbitrates must be active in assessment clinics, i.e. acts as the responsible assessor and must fully participate in MDT meetings" Director 14

6.10.5 Implementation of PHE Arbitration Guidance

The literature review emphasised that producing and disseminating written guidance may have small effects on changes to clinical practice (van Bodegom-Vos et al. 2012). Various factors facilitate or impede guidance use which includes guideline characteristics, professional characteristics and environmental factors (organisational, political, and social factors) (van Bodegom-Vos et al. 2012). To determine the impact of the arbitration guidance respondents were asked if this had/or will change practice in their unit. The results in Graph 14 demonstrate that from a Director's perspective, the guidance has changed practice in only 3 of the respondent units. All three specified they had delegated the role of arbitration to Consultant Radiographers.

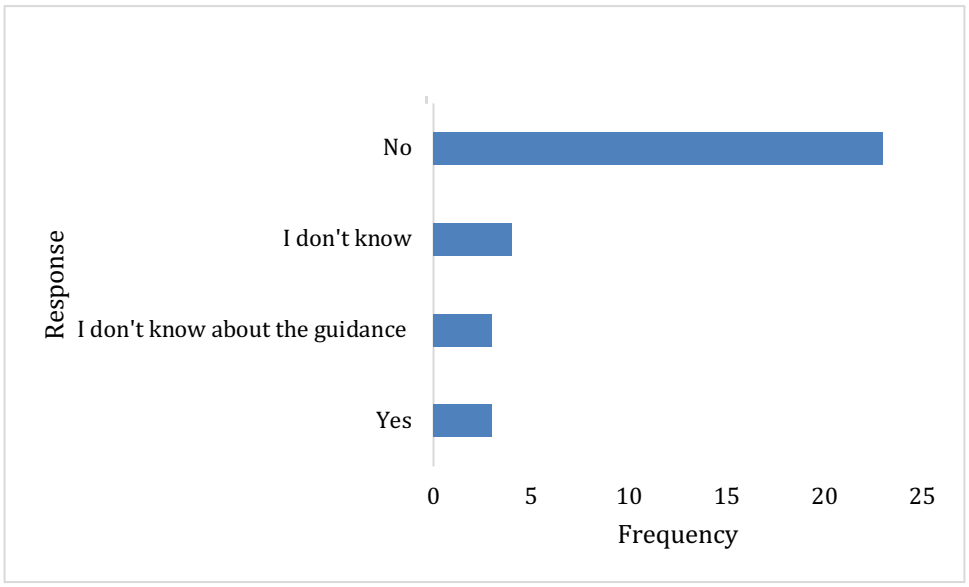


Figure 20 Graph 14. Director Response on Whether PHE Guidance Has or Will Change Practice in Their Unit

Amalgamating data from the Director and film reader surveys produced nine overall non-responses from the 49 units, but the results predominantly still show only a small change in practice (Graph 15).

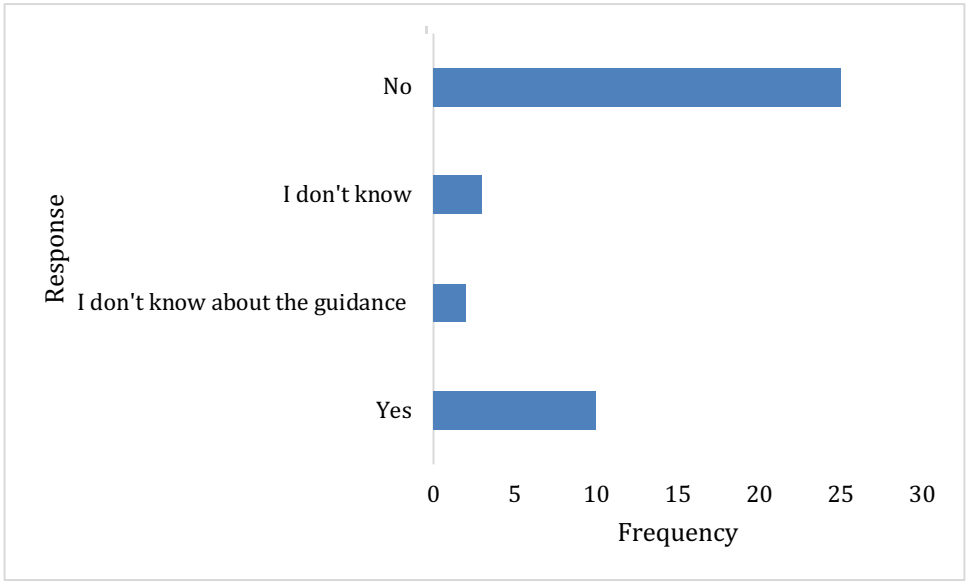


Figure 21 Graph 15. Response From the 49 Units on Whether PHE Guidance Has or Will Change Practice in Their Unit

"Consultant Radiographers were permitted to lead the group instead of there having to be a medic always present" Film reader 42 – Consultant Radiographer

However, it was apparent that there were varying responses to this question from different professional groups working within the same unit. Again, this may reflect that policy has not been disseminated in some units, and there is a variation in understanding and knowledge at different levels of staff.

"We would introduce it if PHE guidance allowed as it would help out our workflow on occasions getting appointments out in a timely manner" Film reader 14 – Advanced Practitioner

6.10.5.1 Potential Barriers to Radiographer Arbitration/Lead of Consensus

In an attempt to understand why Radiographer arbitration (third reader or lead/coordinator of consensus) may not be implemented, non-adopters from both surveys were requested to rate on a 5 point Likert scale how strongly they agreed or disagreed with 11 items (0=strongly disagree, 1=disagree, 2=neither agree nor disagree, 3=agree, 4=strongly agree). The list was non-hierarchical and included clinical performance, cultural, pragmatic and organisational options, and a 'no good reason' statement. Results were aggregated into agreement, neutral and disagreement responses (Figure 22).

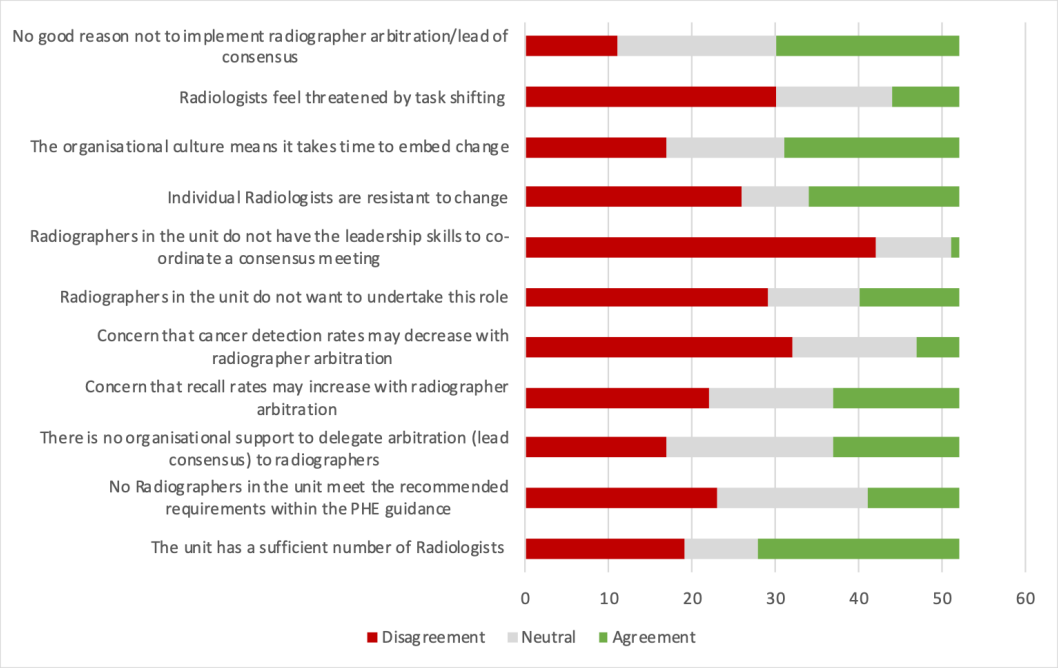


Figure 22 Potential Barriers to Implementing Radiographer Arbitration/Lead of Consensus

Fifty-two respondents (Directors and film readers) completed the above Likert table (Figure 22). The results demonstrate that the main reasons not to implement Radiographer arbitration were units currently had *'sufficient Radiologists'* (46.1%), '*the organisational culture means it takes time to change'* (40.3%), along with there is '*no good reason'* (42.3%) not to implement it. To a slightly lesser extent was '*individual Radiologists being resistant to change'* (34.6%), '*no organisational support to delegate arbitration'* (28.8%) and '*concern that recall rates may increase if undertaken by Radiographers'* (28.8%). 23% agreed with the statement that '*Radiographers do not want to undertake the role'*. A minority (9.6%) agreed that '*cancer detection rates may decrease',* or that '*Radiographers lack the leadership skills to lead/co-ordinate consensus meetings'* (1.9%).

Several free-text comments were made about possible constraints or challenges in

implementing Radiographer arbitration/lead of consensus and elucidated that in one unit it was the recommendation at a prior QA visit (3 years ago) that stipulated a Radiologist should perform this task. In some centres, Radiographers were relatively new to film reading, but there was support to delegate when Consultant Radiographer status was achieved and if there became a departmental need.

"Readers are fairly new hence inexperienced but will eventually acquire Consultant Mammographer status in the years to come. Following which they could lead the arbitration process" Director 20

"This has not been necessary, however, if required in the future there are no team objections to having a Consultant Radiographer lead consensus" Director 21

One Director remarked that they did not have any Consultant Radiographers, and the Advanced Practitioners did not meet the recommended requirements. In their opinion, the guidance was "overly restrictive and has missed the boat" describing that units with staffing shortages had already adopted what they feel is best practice.

"Horses and stable doors" Director 22

21.1% of the respondents were in agreement that Radiographers within their unit did not meet the recommended requirements as specified within the PHE guidance. This was explored further in the survey by specifically routing the Radiographers

(Advanced practitioners and Consultant Radiographers) to a specific sub-section. The results are presented in section 6.10.6 (Figure 23 and Figure 24).

To assess the extent to which responses differed between individual professional roles, a Kruskal-Wallis test was undertaken. This demonstrated significant differences (p=0.013 and p=0.041 respectively) between Consultant Radiographers and Advanced Practitioners in response to '*no organisational support to delegate arbitration'*. The Advanced Practitioners agreed with this statement scoring the highest and Consultant Radiographers disagreed with the statement. Again, Advanced Practitioners perceive that the '*organisational culture means it takes time to change',* but the Consultant Radiographer responses did not support this. These results indicate that Consultant Radiographers have a different perception to Advanced Practitioners regarding task shifting and local organisational culture.

6.10.6 Radiographers Attaining the Recommended Requirements for Delegation of Arbitration/Lead of Consensus

Third reader arbitration/lead of consensus requires individuals to be highly specific and hence reduce needless recalls. This skill is acquired via continuous learning and consequently, why the PHE guidance recommends feedback from decision-making, audit, continuous professional development (CPD) and case review. To determine the extent to which Radiographers perceived they met the recommendations within the guidance, they were asked to rate how often they met the criteria. The results (Figure 23) demonstrate that in the vast majority of cases respondents are annually achieving the requirements of reading sufficient volumes of films (92.3%) (including

first reads, 97.4%), undertaking the PERFORMS test (94.9%) and undergoing an appraisal with a subsequent personal development review (97.4%).

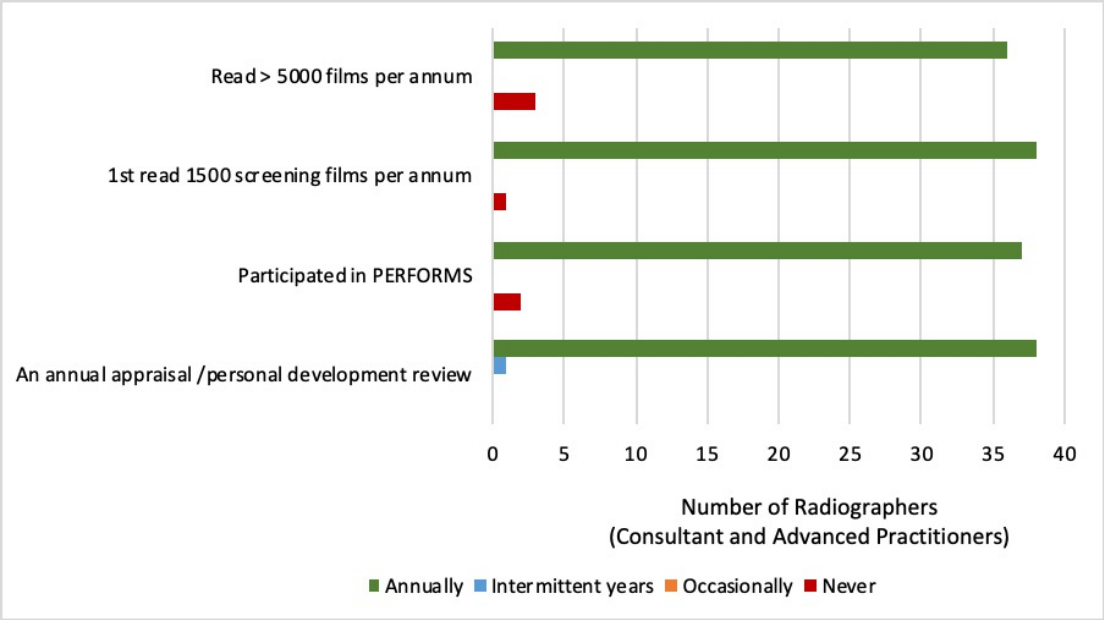


Figure 23 The Frequency of Respondent Radiographers Meeting the Performance Criteria Within the PHE Guidance

However, Figure 24 demonstrates that while the majority undertake a regular audit of their reading practice (97.4%), less are involved in team audit (69.2%). This echoes the findings in section 6.9.10, where team mechanisms to monitor consensus outcomes scored lower. 82.1% stated they evidenced reflective learning from the review of interval cancers, previously assessed intervals and screen-detected cancers. Less than 100% in this response could reflect that reflective learning is taking place, but not necessarily being evidenced. Nevertheless, the categories not being consistently met relate to autonomous decision-making in assessment clinics and actively participating in decision-making and subsequent patient management within MDT'S. Interestingly, cross-tabulation of the results revealed that although those not undertaking the decision-making elements were

mainly Advanced Practitioners, two Consultant Radiographers reported they did not currently meet these criteria. A further two did not undertake regular audit and review of team results or were working towards reflective practice.

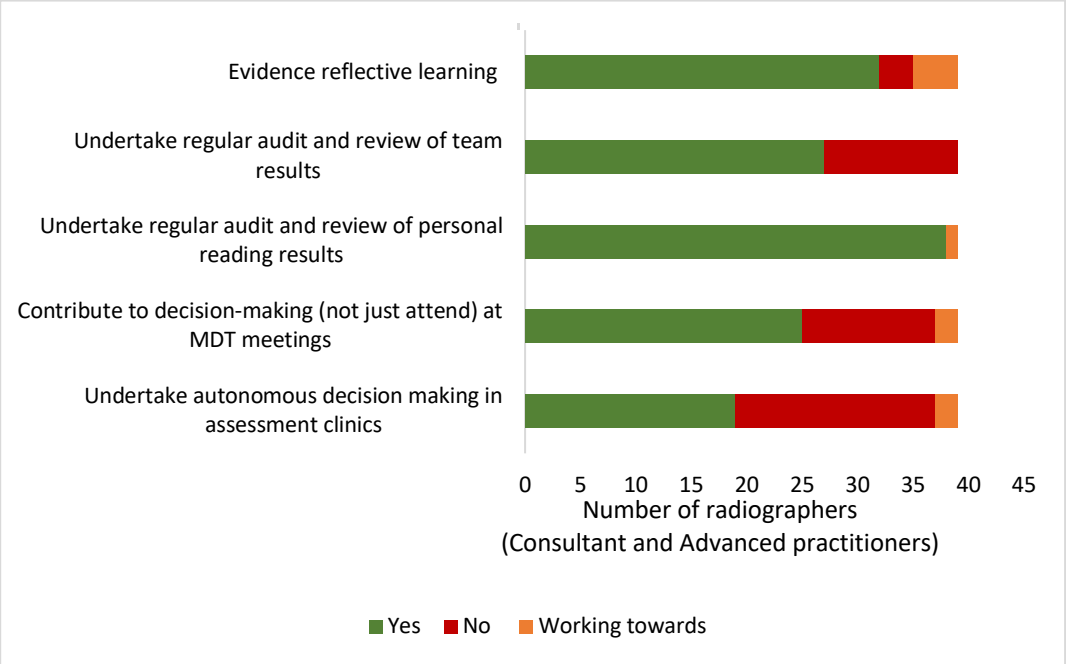


Figure 24 The Number of Respondent Radiographers Attaining the Recommended Requirements for Delegation of Arbitration/Lead of Consensus

While the PHE guidance does not stipulate accreditation with the SCoR is a requirement, it incorporates the statement.

"The Society and College of Radiographers (SCoR) can provide accreditation of Advanced and Consultant Practitioners regarding the four pillars of practice which include: leadership, CPD and education, clinical practice and audit/research capabilities".

The survey result shows that in the main (71.8%) respondent Consultant Radiographers and Advanced Practitioners have never been accredited (Figure 25). The context of the above statement within national guidance was explored further in the subsequent qualitative interview's (Chapter 8).

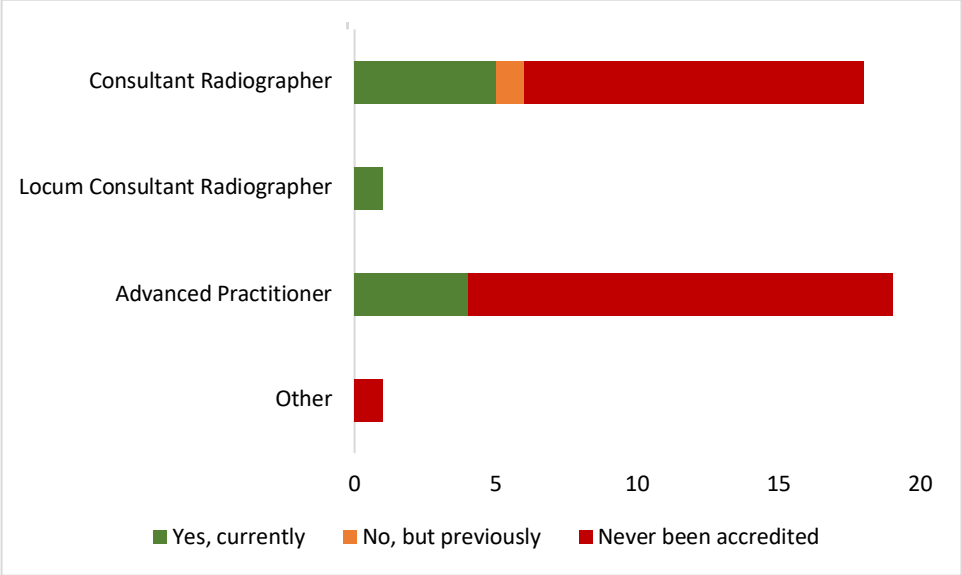


Figure 25 The Number of Respondent Consultant Radiographers and Advanced Practitioners Accredited with the SCoR

6.11 Other Issues Raised by Survey Respondents

One Director raised the critical issue of Radiographers being subjected to a medico-legal claim if they dismissed cancer when acting as the third reader arbitrator

"I have personal concerns about exposing a third read arbitrating Radiographer to medicolegal claims where a cancer is missed" Director 17

This issue was explored with the respondent in the subsequent telephone interview (Chapter 8). However, e-mail correspondence with the professional officer for The Society and College of Radiographers conveyed that they do not have specific data on this. This led to contact with the organisation NHS Resolutions who reported that the highest number of claims where they were asked to identify an expert witness was for missed foetal anomaly examinations (undertaken and reported almost exclusively by sonographers). They stated:

> "*We have had a few breast reporting related cases but have also had cases where physical injury as a result of the mammogram has been alleged"*.

It is inevitable that as Radiographers are increasingly involved with reporting and arbitration of screening cases that the numbers of claims against these staff may increase.

6.12 Chapter Summary

This chapter has discussed the relevance of a national electronic survey in relation to the current study and presented an analysis of the survey responses. The main findings demonstrate that:

- There is national variance in all elements of reporting and arbitration practices.
- From the respondent results, non-blinded reading is predominantly undertaken, which may have repercussions on the decision-making of the second reader, and potential impacts on recall rates.

- The pairing of the reporters is mainly based on a professional role rather than performance measures as suggested by the NHSBSP.
- There is also variation as to which cases (concordant/discordant) are evaluated and the professional roles undertaking/leading the process.
- Staff are influenced by knowing who has recalled a case, and this may bias decision-making at the review.
- There is a lack of clarity of the definition of arbitration and consensus, with the two, interchanged and divergent views on the advantages and disadvantages of the processes.
- The PHE guidance on arbitration has had minimal impact on the units that responded. However, there are several recommended requirements in the guidance that Radiographers (Advanced Practitioners and Consultant Radiographers) are not currently attaining.

The next chapter explains the NHSBSP Central Return Dataset (KC62), and the rationale for the parameters used to compare the performance of units. Recall rates are correlated with cancer detection rates. Survey results on reporting and arbitration strategies are analysed relative to performance metrics.

Chapter 7. KC62 Performance Data and Analysis of Unit Performance Based on Specific Metrics

7.1 Introduction

The NHSBSP Central Return Dataset (KC62) captures statistical data on an annual basis. KC62 data is validated and analysed by NHS Digital to monitor the quality and effectiveness of breast screening. It is also utilised to monitor individual programmes performance regarding the achievement of cancer targets and facilitate comparisons with other units regionally and nationally. The annual KC62 central return dataset provides published data on the numbers and age groups of women invited, the acceptance of screening, and the outcomes from the 80 breast screening units in England.

Breast Screening Information System (BSIS) is a tool that has been created by the screening group of PHE. Compared to the previous regional reports, the BSIS system provides reports which demonstrate how a particular reader has performed compared to all other readers in England. PPV of recalls, cancer detection rates and discrepant cancer detection rate are the metrics used. The benefits of the report for the individual, service and programme are summarised in Table 35.

Table 35. The Benefits of the BSIS Generated Report
(Taken from FRQA Working Group 2017)

Individual reader, the report:
1. Provides a unique, unbiased insight into your reading practice
2. Identifies personal strengths and any weaknesses
3. Provides sequential reports that show a drift in detection rates or other parameters
4. Enables the targeting of personal development to address any possible issues
The service, the report:
1. Provides greater granularity than gross statistics, e.g. recall to assess, standardised detection ratios (SDRs), invasive cancer rates
2. Informs film reading developments and strategies, highlighting instances where protocols should be reviewed and revised
3. Enables targeting of personal development for team members
4. Highlights readers within the team possessing particular strengths that may be useful to support other team members, helping to determine optimal reader pairings
The programme:
1. Informs policy, guidance and, in the future, potentially standards
2. Identifies excellent, specific services from whom all services can learn

7.2 KC62 Data Reviewed for this Study

An advisory group was convened to recommend on the parameters to use to review the performance of units relative to data obtained from the survey. The group consisted of an Assistant Professor of Screening and Test Evaluation (Warwick University), Breast Radiologist and regional QA Radiologist (University Hospital Coventry and Warwickshire), Breast Radiologist (Cambridge University Hospitals) who previously sat on various committees (NHSBSP Radiology QA committee, The Association of Breast Surgeons Audit group, National Evaluation group and the President of the European Society of Breast Imaging) and the National Lead for Screening QA Services at PHE Screening.

Initial discussions involved using CDR, PPV and recall rates. It was advised that overall CDR would be more relevant for future publication purposes, particularly as

the USA do not report SDR. However, on reviewing the KC62 published data tables, PPV and overall CDR are not reported (CDR are split by prevalent and incident). To obtain this information would require an Office for Data Release (ODR) request form to be approved by the PHE Breast Screening Research Advisory Committee (RAC). At that time the RAC was still in the set-up phase with no confirmed meeting dates scheduled. Therefore, this would have delayed the study significantly past the identified timeframe and the decision was made to use the performance criteria that was published and freely available. The published outcome statistics and explanations are provided in Table 36. It is acknowledged that cancer detection rates are affected by the age distribution and the background incidence of the disease in specific catchment areas (Blanks, Wallis, and Moss 1998). It is also recognised that the SDR was not intended as a measure between individual programmes (Blanks, Wallis, and Moss 1998). Incidence rates usually increase with age, and although the KC62 returns provide invasive cancers detected for five-year age bands, this was not split by prevalent and incident screens at the individual unit level.

The chosen performance metrics were therefore limited to recall rates (overall, prevalent and incident), small <15mm CDR (prevalent and incident) and SDR (prevalent and incident) (Appendix 9). These metrics were correlated with arbitration strategies, reading type (blinded vs. non-blinded), units utilising Radiographer arbitration and programme size. The resources (time, number of staff, number of cases, number of sessions) required for consensus meetings were also

analysed relative to overall recall rates and SDR. The results are discussed in section 7.12.

Table 36. The Outcome Statistics and Explanations for Recall Rates, and Cancer Detection Rates Published for Individual Screening Programmes (taken from PHE NHSBSP, 2016-17 Publication)

Measure	Explanation
Overall Assessment Rate	Number of women referred for assessment as a percentage of all women screened
Prevalent Assessment Rate	Number of women referred for assessment as a percentage of all prevalent women screened
Incident Assessment Rate	Number of women referred for assessment as a percentage of all incident women screened
Prevalent – small cancers detected (<15mm) per 1,000	The number of women with invasive cancers smaller than 15mm in diameter detected per 1,000 prevalent women screened
Prevalent Standardised Detection Ratio (SDR)	Prevalent cancers. Measures the ratio of screen-detected invasive cancers divided by the expected number of invasive cancers. Applies criteria from the Swedish Two-County randomised control trial, which is used as a benchmark of performance. An SDR of 1 would imply the observed number of invasive cancers is the same as that expected, greater than 1 would indicate higher, and less than 1 lower than the Swedish Two-county study.
Incident - small cancers detected (<15mm) per 1,000	The number of women with invasive cancers smaller than 15mm in diameter detected per 1,000 incident women screened
Incident Standardised Detection Ratio (SDR)	Incident cancers. The ratio of screen-detected invasive cancers to the expected number of invasive cancers pertaining to criteria from the Swedish Two County randomised control trial. An SDR of 1 with an uptake of 70% should achieve a mortality reduction of approximately 25%.

7.3 Size of Unit

The population covered by screening units varies considerably in size. The Forrest report (Forrest 1986) recommendations regarding target populations (41,150) were based on inviting women 50-64 years, on a three-yearly basis with an assumed uptake of 70%. Based on these figures, and allowing for technical recalls and self-referrals, a screening centre was estimated to have 12,000 attendances per annum. Uptake rates will naturally fluctuate yearly across individual programmes. Programmes that have lower uptakes would, therefore, have to invite from larger target populations to accomplish the same minimum volume comparative to programmes with high uptakes. The smallest unit in 2017-2018 screened 5,586 women aged 50-70 and the largest screened 50,429. Unit size is particularly important when comparing data for prevalent screens as this represents a much smaller number of women. Hence, the comparison between units can be prone to error and less reliable.

Research undertaken by Blanks et al. (2002) reported that the performance of smaller screening programs was inferior to medium and large programmes. The performance was based on the number of cancers detected, recall rate to assessment and positive predictive value (PPV) for assessment. The size of the program was classified by the uptake (number of women attending for prevalent and incident screening) within a calendar year, and categorised into three groups (small, medium, large). Small programmes were classified as the bottom 25%, medium programmes as the middle 50% and a large programme, the top 25%. An explanation for the differences in performance was not evident, but it was

acknowledged that for smaller unit's underperformance may be difficult to identify as there is statistical instability when dealing with relatively small numbers.

Since 2002, an amalgamation of units has occurred, and in 2016/17, there were 80 breast screening units rather than 95, and the NHSBSP routinely invited women aged 50-70. The Blanks et al. (2002) division of small, medium and large was deemed appropriate for the classification of unit size in this study. However, performance data over four years was considered a minimum to account for any peaks or troughs within a single year. The size of the program was classified by the average uptake (number of women attending for prevalent and incident screening) from 1 April 2013 to 31st March 2017. Data collated was limited to the screening age 50-70 that all programmes routinely invited as the age extension (47-73) has been phased in at different time intervals across units.

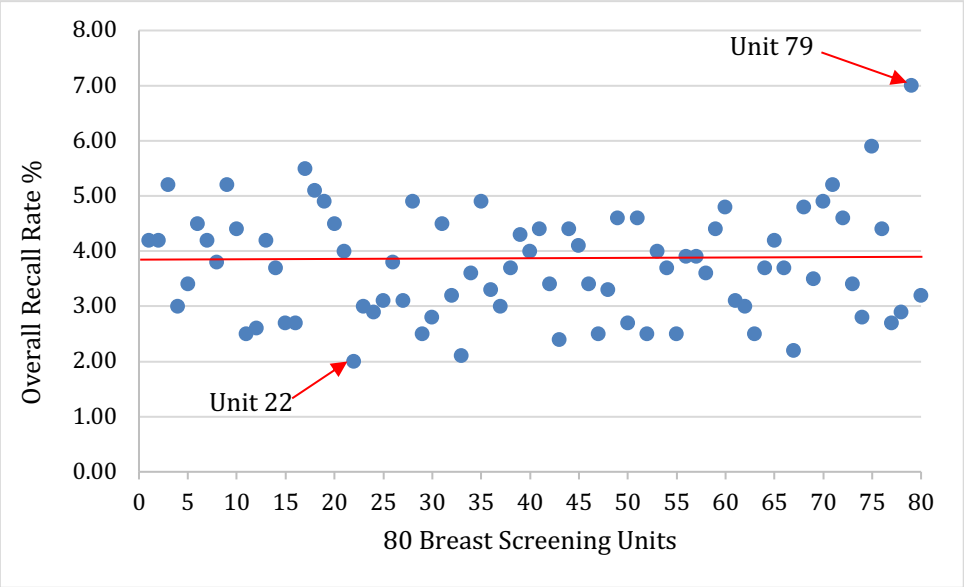
Based on this data with 80 programmes, there are 20 small programmes, 40 medium, and 20 large. Following this principle, a small programme was classified as one with a total annual screening attendance of 6,659 -14,726, a medium size programme was 14,929 -30,226 and a large programme 30,277 – 50,224. The median number of women screened is detailed in Table 37.

Table 37. The Median and Range of Women Screened by Programme Size

Size of unit (no of programmes)	**Median(range) of programme sizes (2013-2017 data)**
Small (20)	10,939 (6,459 -14,726)
Medium (40)	21,401 (14,929 -30,226)
Large (20)	35,929 (30,277– 50,224)

7.4 Overall Recall Rates

As discussed in Chapter 2, there is national variance in recall rates, and third reader arbitration or consensus group review can be fundamental in achieving the national standards. Graph 16 demonstrates the current overall recall rates for the individual 80 breast screening units, which range from 2.04% to 7.04% with a mean of 3.80% (statistics for 2016-2017).

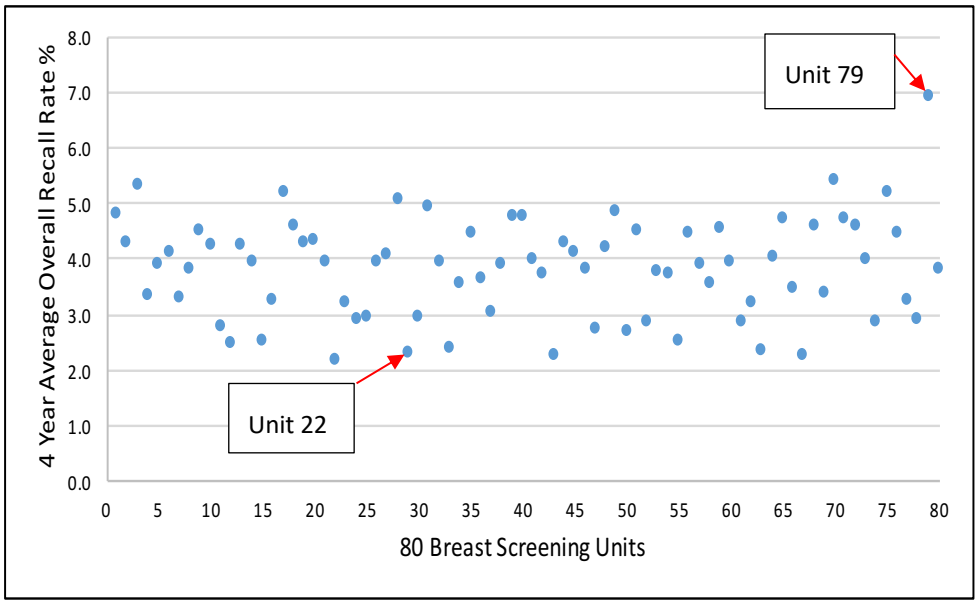


*Red line=mean of 3.80%

Figure 26 Graph 16. Demonstrating the Overall Recall Rates (2016-2017) for all 80 Breast Screening Units in England

To account for yearly variation, the data was also reviewed for the last four years. This demonstrates that there is little difference in the 4-year average overall recall rates ranging from 2.14% to 6.92% with the same mean of 3.80%, as shown in Graph 17. The data also demonstrate that units are relatively constant in their recall rates, the positions of most units on the graph being consistent. Units with lower overall recall rates in 2016-2017 had similar 4-year average rates, e.g. unit 22, 2016-2017

recall rate of 2.0% and a 4- year average of 2.14%. Similarly, units with high overall recall rates (unit 79, 2016-2017 7.0%) were consistently high over the four years (6.9%).

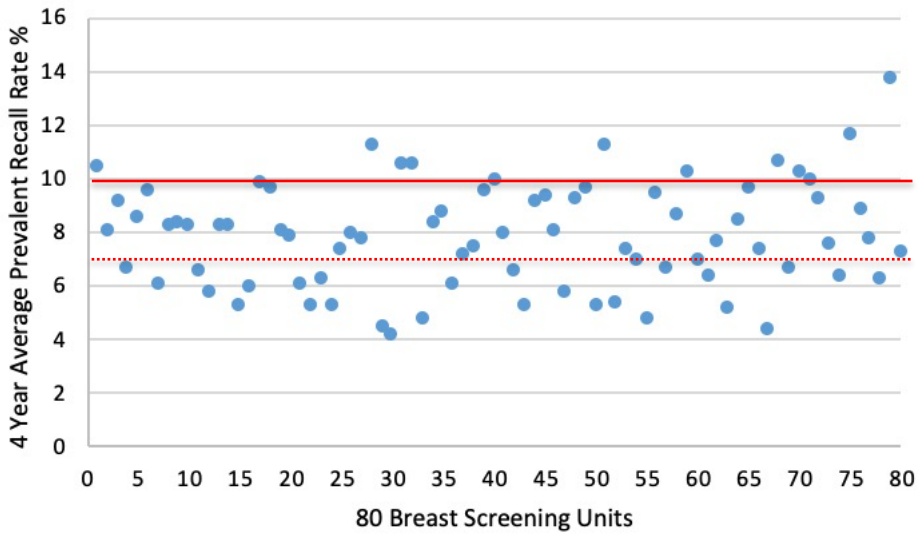


*Red line=mean of 3.80 over four years

Figure 27 Graph 17. Demonstrating the 4 Year Average Overall Recall Rates (2013-2017) for all 80 Breast Screening Units in England.

7.5 Prevalent and Incident Recall Rates

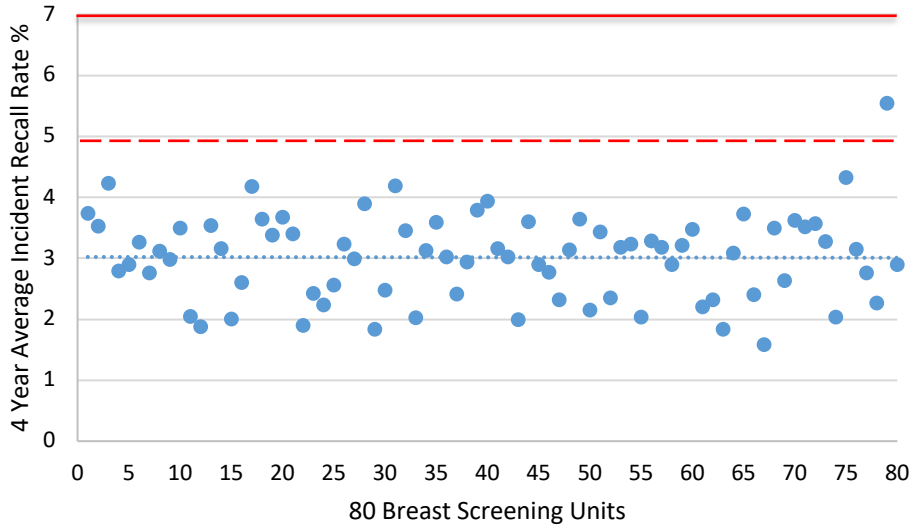
Four-year (2013-2017) data for prevalent and incident recall rates were also reviewed separately. Graph 18 demonstrates that prevalent recall rates ranged from 4.2% to 13.7% with ten units failing to meet the NHSBSP acceptable threshold of <10% and fifty-one units are above the achievable threshold of <7%.



Solid red line=Acceptable threshold of 10%, Dashed red line=Achievable standard of 7%

Figure 28 Graph 18. Demonstrating the 4 Year Average Prevalent Recall Rates (2013-2017) for all 80 Breast Screening Units in England.

For incident screens, Graph 19 demonstrates that incident recall rates ranged from 1.6% to 5.5% with all units below the NHSBSP acceptable threshold of <7% and only one unit above the achievable threshold of <5%.



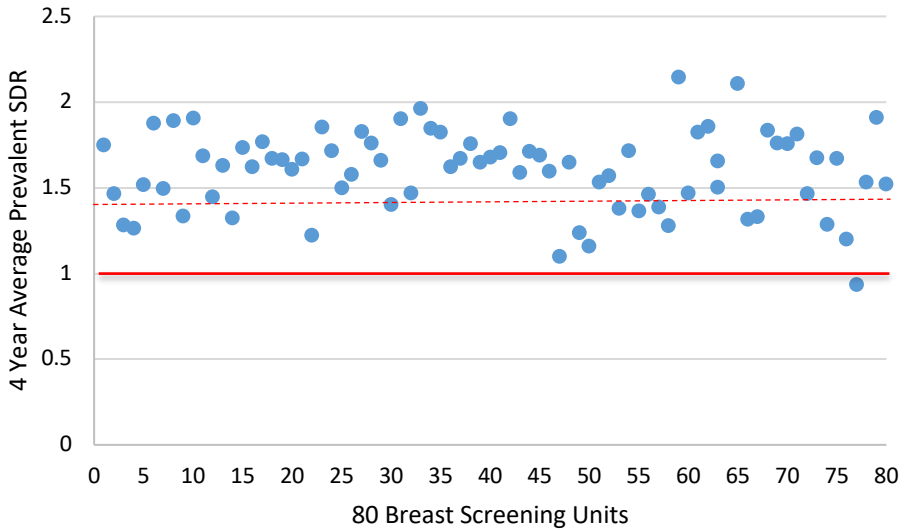
Solid red line=Acceptable threshold of 7%, Dashed red line=Achievable standard of 5%

Figure 29 Graph 19. Demonstrating the 4 Year Average Incident Recall Rates (2013-2017) for all 80 Breast Screening Units in England.

This demonstrates that units are struggling to achieve the NHSBSP recall targets for prevalent screens but not for incident screens. However, the impact of over recalling in the incident screens is more significant as there are four times as many clients (53-70 age).

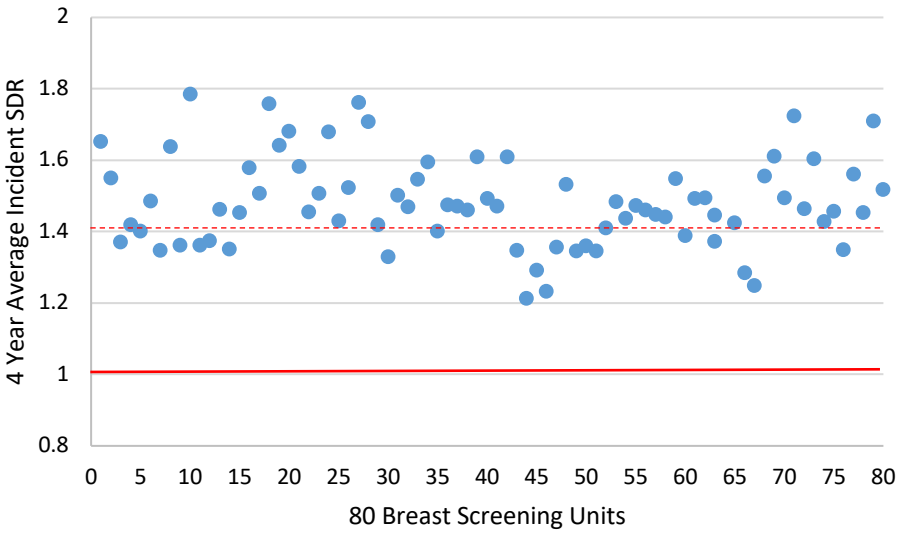
7.6 Prevalent and Incident SDR (2013-2017)

KC62 data for the same period (2013-2017) was reviewed for prevalent and incident SDR. The NHSBSP minimum standard SDR is ≥ 1.00, with a target of ≥ 1.4 for both prevalent and incident screens. Graph 20 shows that during this period for prevalent screens, only one unit fails to meet the minimum standard, but an additional sixteen fail to meet the target for prevalent SDR. Graph 21 shows the results for incident screens. All units achieved the minimum standard for incident SDR, but twenty fail to meet the target.



Solid red line=minimum standard of ≥ 1.00, Dashed red line= target SDR of ≥ 1.4

Figure 30 Graph 20. Demonstrating the 4 Year Average Prevalent SDR (2013-2017) for all 80 Breast Screening Units in England.



Solid red line=minimum standard of ≥ 1.00, Dashed red line= target SDR of ≥ 1.4

Figure 31 Graph 21. Demonstrating the 4 Year Average Incident SDR (2013-2017) for all 80 Breast Screening Units in England.

Based on the KC62 SDR results for the past four years, the SDR for both prevalent and incident screens showed considerable variability in some units. When the 80 units were ranked in ascending order, a unit that was at the bottom in 2015-2016 with an incident SDR of 1.04 was near the top in 2016-2017 with an SDR of 1.95. A different unit with a low prevalent SDR in 2015-2016 ranked fourth moved to one of the highest (ranked 78th) for prevalent SDR in 2016-2017. Thus, performance based on the average SDR for four years is difficult with data influenced by the extreme values in single years.

7.7 Correlations

7.7.1 Four Year Average Prevalent Recall rates and Prevalent <15mm CDR's (2013-2017)

The literature review highlighted the international variance in recall rates and the difficulty in achieving a balance between high detection (sensitivity) while limiting excessive false-positive recalls. There is controversy as to whether there is a strong

correlation between recall rates and cancer detection (Gur et al. 2004, Yankaskas et al. 2001b, Otten et al. 2005, Grabler et al. 2017). A recent observational study (Burnside et al. 2018) has analysed screening data from 11,258,620 women in England aged 45-70, between April 2009 and March 2016. The authors concluded that 99% of invasive cancers and high-grade DCIS were detected at an estimated recall rate of 7%. Above this level low and intermediate-grade DCIS continues to be detected, but this is associated with rapid increases in false-positives.

Data were analysed for all 80 breast units in England. A bivariate correlation coefficient was used to determine if there was a correlation between the two variables of prevalent recall rates and prevalent small (<15mm) cancer detection rates. To account for yearly variation, the data was analysed for the four years between 2013-2017. Inspection of the scatterplot of the two variables identified one outlier unit with a high recall rate. Pearson's correlation coefficient, *r*, is sensitive to outliers (Pernet, Wilcox, and Rousselet 2012), which can lead to a value that is an inaccurate summary of the data as a whole. Therefore, the analysis was also conducted with the outlier removed. Both variables were normally distributed, as assessed by Shapiro-Wilk's test ($p > .05$) (with the outlier removed).

The results demonstrate there was no statistically significant correlation between the 4-year average prevalent recall rates and small (<15mm) cancer detection rates regardless of whether the outlier was included or excluded r (77) = 0.220, p = 0.051 (outlier excluded); r (78) = 0.205, p = 0.068 (with outlier included). Although the p-value is close to significance with the outlier excluded, it would only represent a

weak correlation (0.220). Prevalent recall rates explained 4.8% of the variation in small cancer detection rates.

7.7.2 Four Year Average Prevalent Recall rates and Prevalent SDR's (2013-2017)

Data for the same period (2013-2017) was analysed for prevalent recall rates and prevalent SDR's. Again, the analysis was conducted with and without the outlier removed. With the outlier removed, both variables were normally distributed, as assessed by Shapiro-Wilk's test ($p > .05$). The results demonstrate that there is a statistically significant, weak positive correlation between the prevalent recall rate and prevalent SDR, r (77) = 0.255, p=0.023 (outlier excluded) r (78) = 0.288, p = 0.01 (outlier included). However, it is recognised that analysis and comparison between units can be prone to error and less reliable in smaller units as the numbers of cases associated with prevalent data can be small.

7.7.3 Four-year Average Incident Recall rates and Incident <15mm CDR's (2013-2017)

A bivariate correlation coefficient was also used to determine if there was a correlation between the two variables of incident recall rates and incident small (<15mm) cancer detection rates. Both variables were normally distributed, as assessed by Shapiro-Wilk's test ($p > .05$) (outlier removed). As with the prevalent results, there was no evidence to suggest a statistically significant correlation between the 4-year average incident recall rates and small <15mm CDR, r (77), =0.098, p = 0.39 (outlier excluded) r (78), 0.095, p=0.42 outlier included.

7.7.4 Four-Year Average Incident Recall rates and Incident SDR's (2013-2017)

Data for the same period was analysed for incident recall rates and incident SDR. As with the prevalent data, the analysis was conducted with and without the outlier removed. With the outlier removed, both variables were normally distributed, as assessed by Shapiro-Wilk's test ($p > .05$). The results of Pearson's correlation coefficient show that there is a statistically significant, moderate positive correlation between the incident recall rate and incident SDR, r (77), = 0.306, p = 0.006 (outlier excluded) r (78), =0.357, p = 0.001 (outlier included) as demonstrated in Graph 22.

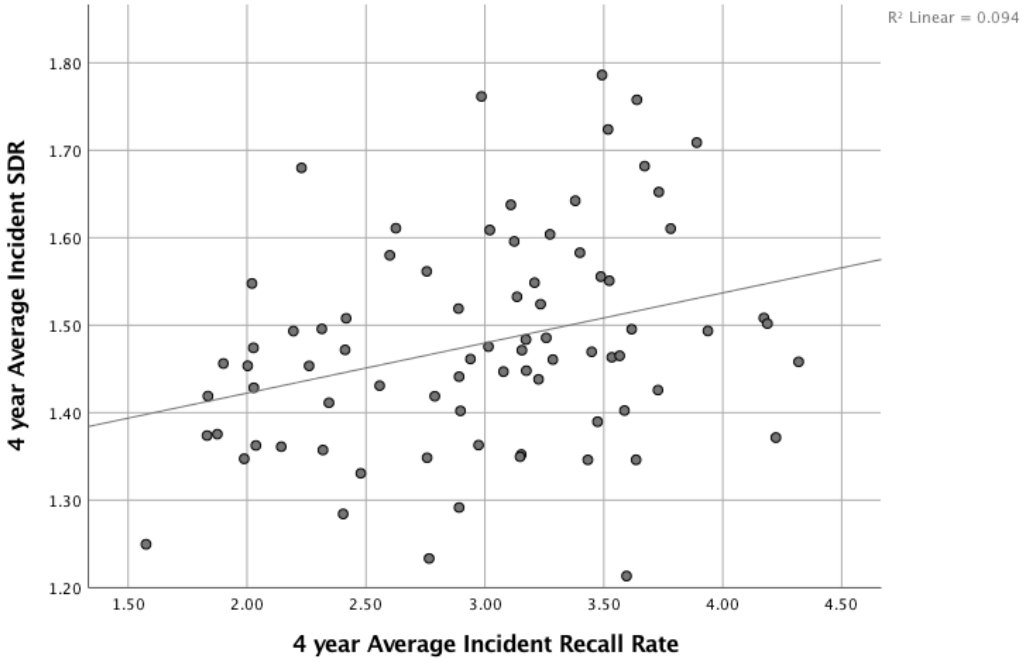


Figure 32 Graph 22. Scatter Plot Demonstrating the Moderate Positive Linear Relationship of 4 Year Average Incident recall rates and Incident SDR (outlier excluded)

Visual inspection of the scatter plot shows that incident SDR increases as a function of recall rates but, above approximately 4% the SDR decreases and so the strength

of the association between recall rates and SDR depends upon how high the recall rate is. A quadratic model was used to predict the maximum recall value on the current data. Graph 23 demonstrates the linear and quadratic trend. For the linear trend, the F-ratio is 7.949, and this is significant at the 0.006 level. For the quadratic trend, the F-ratio is 4.601, significant at the 0.013 level. With a quadratic trend, the peak SDR in this data occurs with a recall rate of 3.781%.

The form of this relationship is

$$Y=ax^2 + bx + c$$

Where x= the recall rate, y= the incident SDR, *a=-0.032 and b=0.242*

'A quadratic expression is a parabola, so it has either a maximum value or a minimum value' (Khan Academy 2019).

and therefore, as the b2 is negative the max value is at $\frac{-b1}{2\ x\ b2}$ *b1= 0.242, b2= -0.032*

-0.242/ 2 x -0.032 = 3.781%

These findings support the NHSBSP achievable standard of <5% for incident screens and are also in keeping with the 2018 (Burnside et al.) study which states that

'a quarter of English screening units currently have recall rates below 2.6%.'

and suggests that a recall rate of approximately 3.1% for incident screens would improve the detection of invasive cancers.

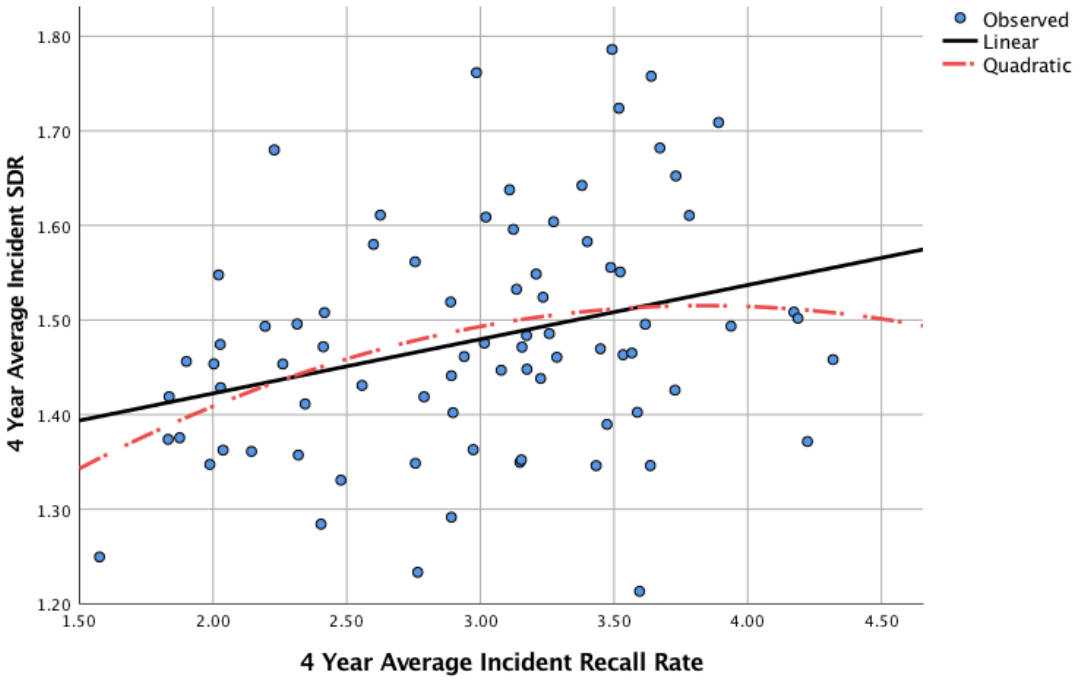


Figure 33 Graph 23. Demonstrating the 4 Year Average Incident recall rates and Incident SDR with a Linear and Quadratic Trend (outlier excluded)

7.8 ANOVA

7.8.1 Prevalent Cases Reviewed and Recall Rates

In the survey, some units responded that the cases arbitrated/reviewed at consensus differed depending on whether it was a prevalent or incident case. Therefore, the strategies were analysed separately. ANOVA tests the null hypothesis that all group means are equal and produces an F-statistic or F-ratio. This was, therefore considered the appropriate statistical test to use. A one-way ANOVA was conducted to determine if the prevalent recall rates were different for the varying strategies. Strategies from the survey were classified into four groups: review discordant cases only (*n* = 27), review all recalls (concordant and discordant) (*n* = 18), other (*n* = 4) and unknowns from non-responders (*n* = 31). The other category

included units specifying that they reviewed discordant recalls and technical recalls (even if both readers agreed in the technical recall). Also, there was a variation in one unit using a third reader to arbitrate all concordant recalls. If the outcome of the case was likely to be benign, that case also went for consensus review. If the features were considered suspicious, the third reader would ratify the recall to assessment independently.

Data for 2013-2017 was analysed. There was one outlier unit in the unknown category as assessed by inspection of a boxplot. Although one-way ANOVA is reported to be rather robust to deviations from normality and does not significantly affect Type I error rates (Maxwell, Delaney, and Kelley 2017), the data were analysed with and without the outlier included (Weisberg 2014). With the outlier excluded data was normally distributed for each group, as assessed by Shapiro-Wilk test ($p > .05$); but the Levene's test for the mean (p = <.05) showed there was heterogeneity of variances. Field (2009 pg: 152) describe homogeneity of variance as:

'The assumption that the spread of scores is roughly equal in different groups of cases'

ANOVA is considered robust when sample sizes are equal, but the accuracy of the F statistic is affected if group sizes are unequal (Wilcox 2012). In this situation, ANOVA is not robust to violations of homogeneity of variance. Not accounting for homogeneity of variance can result in conservative F-ratios producing a result that is non-significant when an actual difference exists in the population. Conversely, a

significant result may be produced when there is no difference between the groups (the Type I error rate is not controlled) (Glass, Peckham, and Sanders 1972).

ANOVA automatically provides a robust ANOVA (Welch's ANOVA) if homogeneity of variance is violated. As there were no specific hypotheses about the effect the strategy might have on recall rates, post hoc tests (Tukey and Games–Howell) were selected (Field 2009) rather than custom contrasts. Tukey is recommended (Westfall et al. 2011, Kirk 2013) when homogeneity of variances is not violated. However, this test is designed for equal numbers in the groups of the independent variable. In this study, there were unequal numbers in the groups, but SPSS Statistics automatically runs the Tukey-Kramer post hoc test if group sample sizes are different (Hayter 1984).

Data are presented as mean ± standard deviation. Recall rates were highest in units reviewing all recalls (8.4 ± 2.0), then for the other strategy (8.0 ± 3.5), discordant only cases (7.8 ± 1.7) and lowest (7.5 ± 2.0) for the unknown (7.3 ± 1.6 outlier excluded), in that order, but the differences between the strategies was not statistically significant, $F(3, 12.777) = 1.444$, $p = 0.276$ (outlier excluded) and Welch's $F(3, 12.894) = 0.808$, $p = 0.512$ (outlier included).

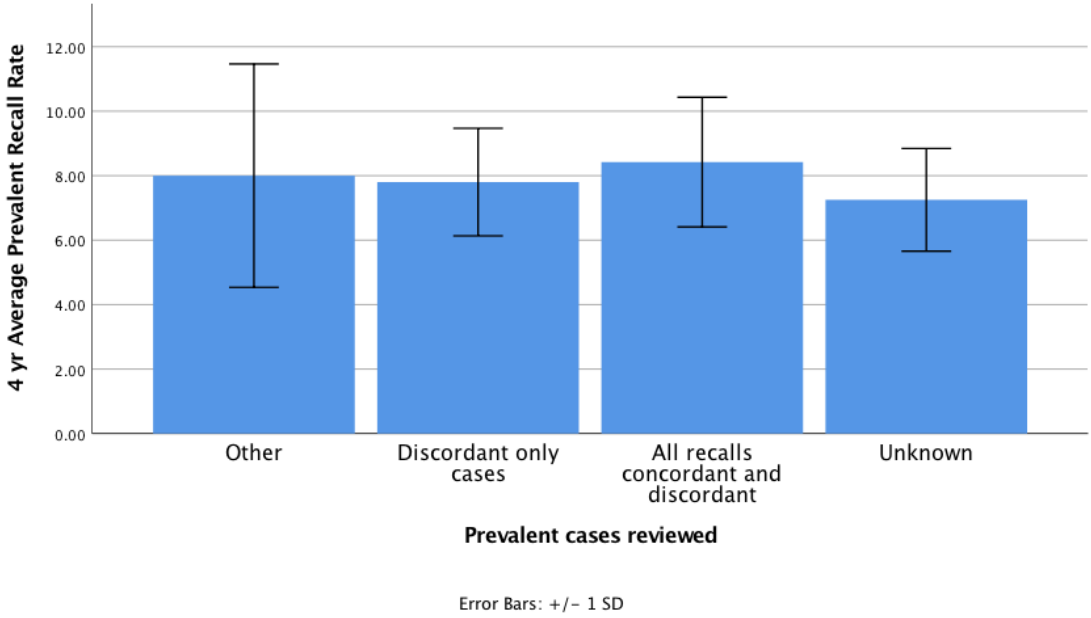


Figure 34. Histogram showing the Mean +/- SD of 4yr Average Prevalent Recall Rates with Cases Reviewed (outlier excluded)

The results in Figure 34 indicate that the strategies are operating the same, i.e. there was no difference in mean prevalent recall rates for those units reviewing all recalls.

7.8.2 Incident Cases Reviewed and Recall Rates

A one-way ANOVA was also conducted to determine if the incident recall rates were different for the varying strategies. As before, strategies were classified into four groups: discordant cases only (n = 32), all recalled cases (n = 15), other (n = 2) and unknowns from non-responders (n = 31). Data for the 4-year average was analysed (2013-2017). The same unit was an outlier in the data as assessed by inspection of a boxplot. As before, this meant the data was not normally distributed for the unknown category, as assessed by Shapiro-Wilk test (p <.05). The one outlier was removed, and the analysis repeated. Variances were homogeneous, as assessed by Levene's test for equality of variances p=.440.

Data are presented as mean ± standard deviation. Recall rates were highest for the discordant only cases (3.1 ± 0.7), all recalls (3.1 ± 0.7), unknown units (2.8 ± 0.6), and lowest for the other strategy (2.7± 0.4). Again, there is no significant difference in the mean incident recall rate between the different strategies regardless of whether the outlier was removed F (3, 75) = 0.917, p=0.437 or included F (3, 76) = 0.388, p=0.762 as demonstrated in Figure 35.

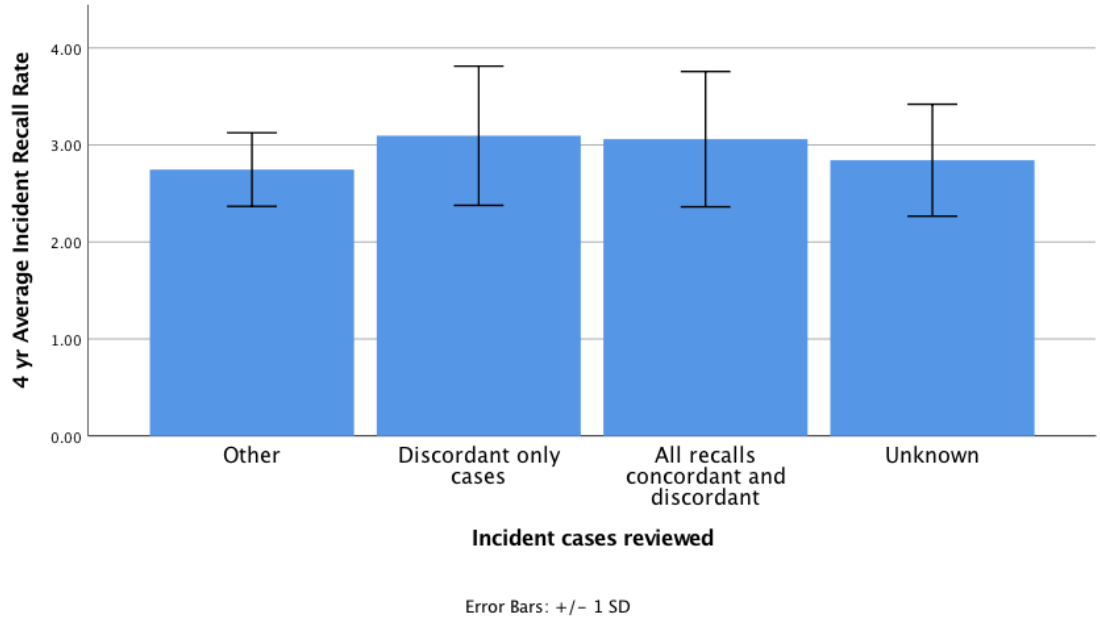


Figure 35. Histogram showing the Mean +/- SD of 4yr Average Incident Recall Rates with Cases Reviewed (1 Outlier removed)

The suggestion to review all recalls is often made at Quality Assurance visits to units with high recall rates, particularly in the prevalent round. However, it is interesting that although not statistically significant, the highest recall rate in the prevalent screens was units reviewing all recalls (concordant and discordant). Without knowing if a significant amount of concordant recalls were returned to routine screening following a review, it is not possible to ascertain if reviewing all recalls is

worthwhile. It may be that a significant proportion of cases are considered benign/normal as per some of the survey comments and that this process brings those unit's recall rates on par with units only reviewing discordant cases. However, survey comments also highlighted that although there were initial improvements in recall rates (decreasing), the benefit was no longer evident and reviewing all recalls was considered rather wasteful of limited time resources. Is the root of the problem actively improving individual film reader recall rates? As discussed in the literature review correlations between recall rates and cancer detection rates are complex, and therefore these are analysed further by mapping to the arbitration strategies in section 7.9.

7.8.3 Arbitration Type and Four Year Average Overall Recall Rates

The survey results demonstrated variance in the arbitration strategies used, some units using both a third reader first –line and if the decision is to recall, then the case is reviewed at a consensus meeting. One-way ANOVA was conducted to determine if the overall recall rates for the same 4-year period (2013-2017) were different relative to the type of arbitration used. Strategies from the survey were classified into four groups: a single third reader (n = 10), group consensus (n=32), mixed strategies (n=7) and the unknowns (n=31). As before the same one unit was an outlier as assessed by inspection of a boxplot for values greater than 1.5 box-lengths from the edge of the box. The analysis was conducted with and without the outlier removed.

With the outlier removed, data was normally distributed for each group, as assessed by the Shapiro-Wilk test (p >.05). Variances were homogeneous, as assessed by

Levene's test for equality of variances (p >0.05). Data are presented as mean ± standard deviation (Figure 36). Overall recall rates increased from the unknown group (3.6 ± 0.8), to the consensus group review (3.7 ± 0.9), to single third reader (4.0 ± 0.8) to the mixed strategies (4.5 ± 0.5) groups, but the differences between the arbitration groups were not statistically significant F (3, 75) = 2.589, p = 0.059 (outlier excluded), F (3, 76) = 1.801, p = 0.154 (outlier included).

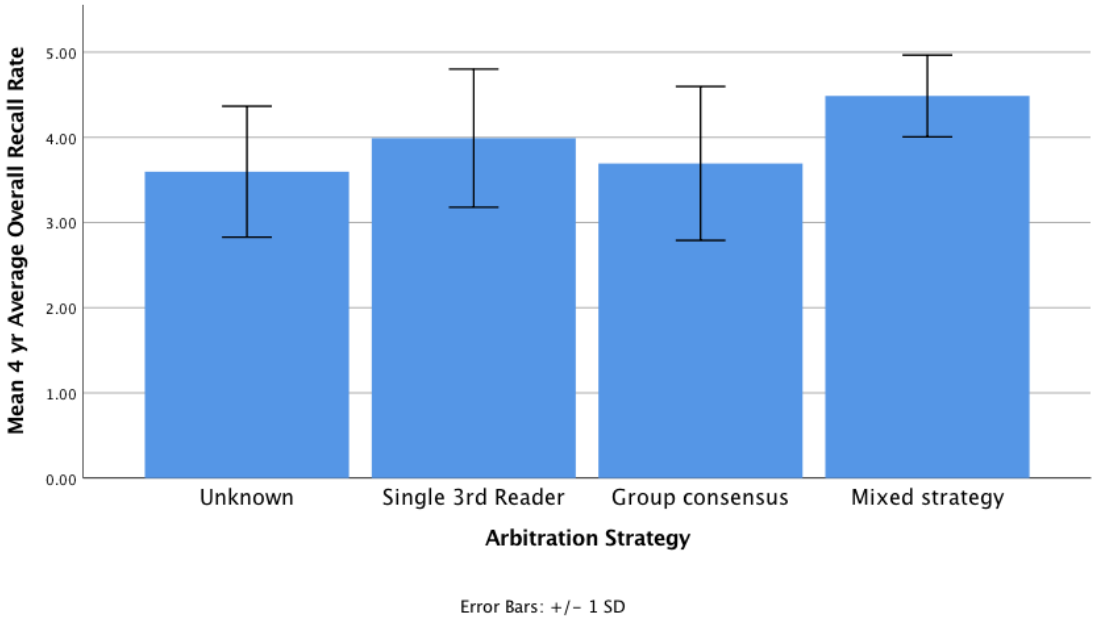


Figure 36. Histogram showing the Mean +/- SD of 4yr Average Overall Recall Rate Relative to the Arbitration Strategy

7.8.4 Reading Type (blinded, non-blinded, partially blinded) and Four Year Average Overall Recall Rates

Free text comments from the survey highlighted that some respondents felt they were influenced by reading practices and the ability to see what the first reader had reported. This is particularly important if units are reviewing only the discordant cases as a second reader may be influenced to recall based on the first reader's decision. Therefore, the data was first analysed with one-way ANOVA to determine

if the reading type demonstrated a statistically significant difference on average overall recall rates (2013-2017). Reading types were classified into three groups: blind reading (n = 4), partially blinded (n=14) and non-blinded (n=31) and the unknowns (n=31). Data was again analysed with and without the outlier unit.

Data were normally distributed for each group, as assessed by the Shapiro-Wilk test (p >.05) with the outlier excluded. Homogeneity of variance was met (Levene's test p = .286). Data are presented as mean ± standard deviation (Figure 37). Overall recall rates increased from the partial blinding (3.5 ± 0.9), to the unknowns (3.6 ± 0.8), to fully blinded (4.0 ± 0.5) to the non-blinded units (4.0 ± 0.9) groups, but the differences between the reading types were not statistically significant, F (3, 75) = 1.984, p = 0.124 (outlier excluded) F (3, 76) =1.309, p=0.278 (outlier included).

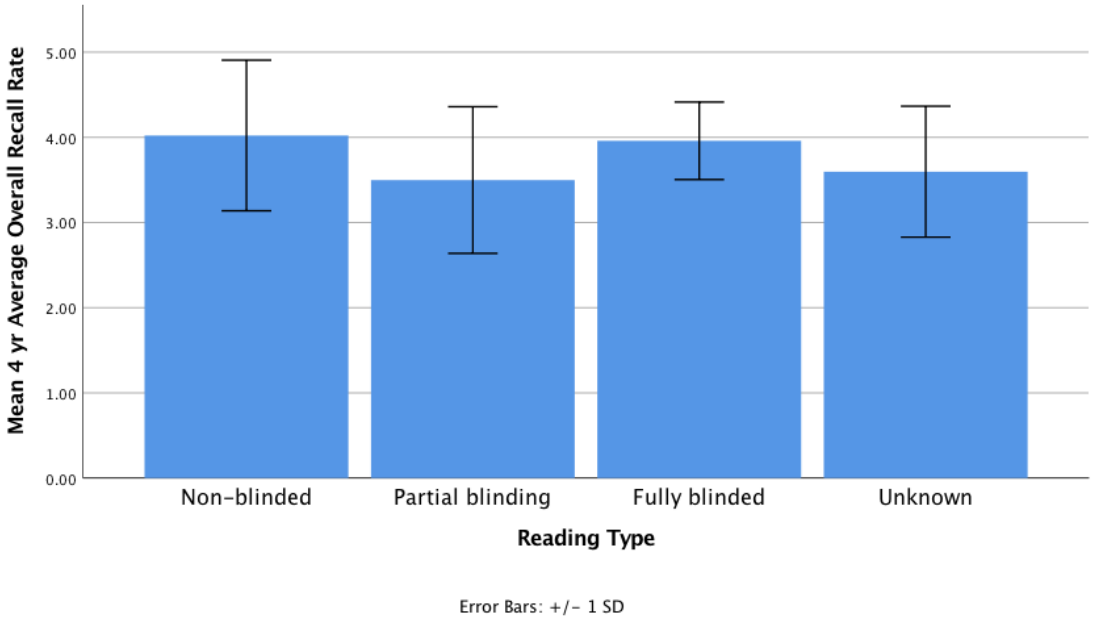


Figure 37. Histogram showing the Mean +/- SD of 4-Year Average Overall Recall Rate Relative to the Reading Type.

7.9 Two-way ANOVA

A two-way ANOVA tests for differences in the effects of two nominal independent variables on a dependent continuous outcome variable. Therefore, to examine the effects of the reading type (blinded, partially blinded and non-blinded) and the cases reviewed (discordant only vs. all recalls) on prevalent overall recall rates (2013-2017) a two-way ANOVA was conducted.

7.9.1 Prevalent Reading Type, Cases Arbitrated and Four Year Average Overall Recall Rates

Residual analysis was performed to test for the assumptions of the two-way ANOVA. Outliers were assessed by inspection of a boxplot; normality was assessed using Shapiro-Wilk's normality test, and homogeneity of variances was assessed by Levene's test. There was one outlier being greater than 1.5 box-lengths from the edge of the box in a boxplot. The one outlier was therefore removed, and the analysis repeated. Residuals were normally distributed ($p > .05$). The assumption of homogeneity of variances was violated, as assessed by Levene's test for equality of variances, $p = .023$ based on the mean value. Two-way ANOVA is also stated to be rather robust to heterogeneity of variance (Jaccard 1998), and therefore the test was still considered appropriate.

Visual inspection of the profile plot gave an initial impression of an interaction between the two independent variables. However, as stated by Fox (1991), profile plots cannot determine an interaction effect as they are based on *sample data* and may reflect a sampling error. Therefore, statistical significance testing is required to

test for the presence of an interaction effect. These results show the interaction effect between reading type and prevalent cases reviewed on overall prevalent recall rates was not statistically significant with F (3, 71) = 0.537, p = 0.659, partial η^2 = .022 or without the outlier unit F (3, 70) = 0.624, p = 0.602, partial η^2 = .026 as demonstrated. The profile plots were then reproduced with standard error (SE) bars (Figure 38,39) which demonstrated large error values.

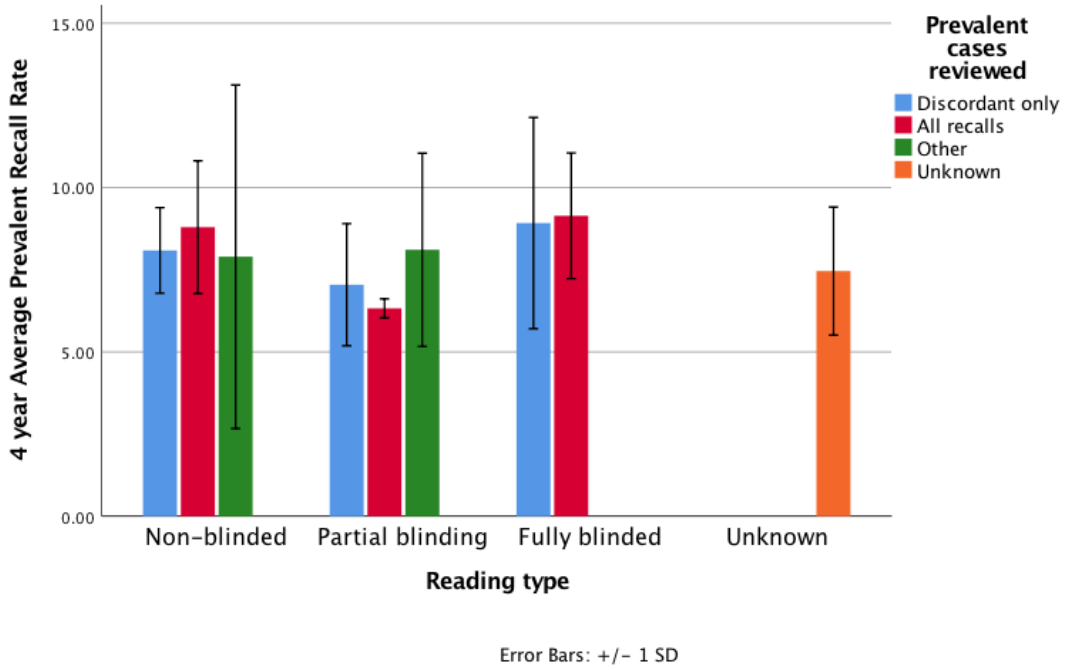


Figure 38. Clustered Bar Mean of 4-Year Average Prevalent recall, Reading Type and Prevalent Cases Reviewed.

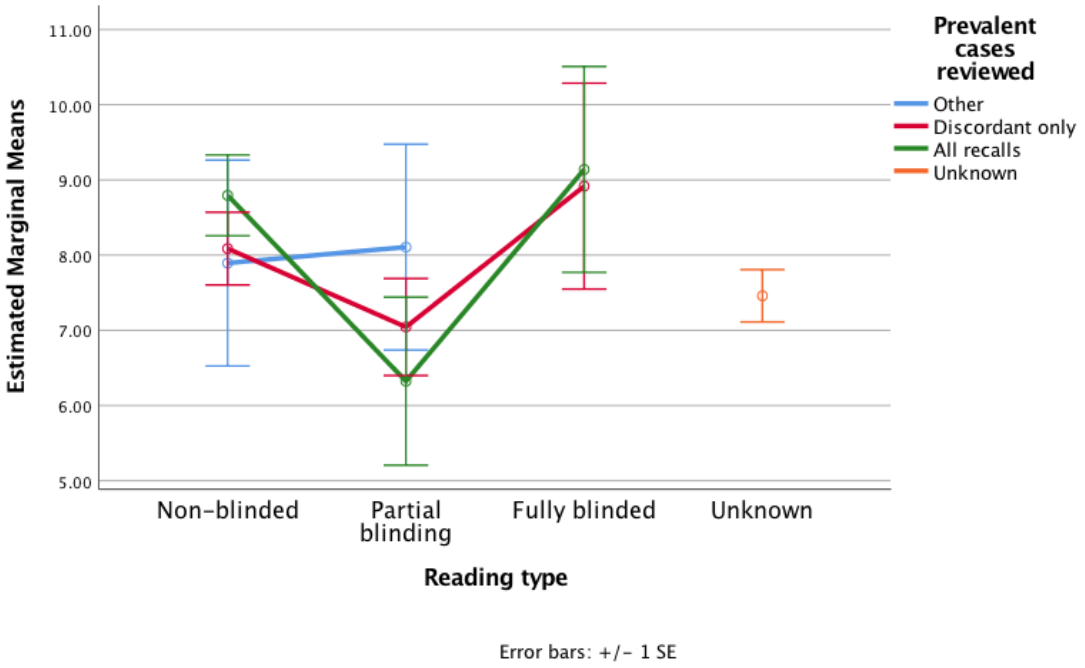


Figure 39. Estimated Marginal Means of 4-Year Average Prevalent Recall Rates

When there is no statistically significant interaction effect, it is recommended to analyse the main effect (Faraway 2014). Searle (2006) reports that a

'non-statistically significant interaction effect is not evidence of its absence (i.e., not rejecting the null hypothesis does not mean accepting the null hypothesis)'

This was performed for reading type and indicated there was no statistically significant main effect of reading type, or cases reviewed on prevalent recall rates (with or without the outlier included) as demonstrated in Table 38.

Table 38. Test of Between-Subject Effects. Dependent Variable 4year average prevalent recall rate

Outlier Included					Outlier Excluded				
Source	df	F	Sig.	Partial Eta Squared	Source	df	F	Sig.	Partial Eta Squared
Reading type	2	1.725	.186	.046	Reading type	2	2.005	.142	.054
Cases reviewed	2	.097	.908	.003	Cases reviewed	2	.0113	.893	.003
Error	71				Error	70			

The 'unknowns' were also removed from the dataset and the procedure repeated. The results (Table 39) still show no main effect for reading type or cases reviewed on the four-year average prevalent recall rate.

Table 39. Test of Between-Subject Effects. Unknowns removed. Dependent Variable 4year average prevalent recall rate

Source	df	F	Sig.	Partial Eta Squared
Reading type	2	1.742	.188	.078
Cases reviewed	2	.098	.907	.005
Error	41			
Total	49			
Corrected Total	48			

7.9.2 Incident Reading Type, Cases Arbitrated and Four Year Average Overall Recall Rates

A two-way ANOVA was also conducted to examine the effects of the reading type (blinded, partially blinded and non-blinded) and the cases reviewed (discordant only vs. all recalls) on incident overall recall rates (2013-2017). Again, there was the same outlier unit identified, and therefore the analysis was repeated with the outlier removed. Residuals were normally distributed ($p >$ .05). Variances were

homogeneous, as assessed by Levene's test for equality of variances, p = .109. The interaction effect between reading type and incident cases reviewed on overall incident recall rates was not statistically significant, F (3, 71) = 0.410, p = 0.746, partial η^2 = .017 (with outlier) with or without the outlier unit F (3, 70) = 0.498, p = 0.685, partial η^2 = .021 (outlier removed). The profile plots with standard error (SE) bars (Figure 40, 41) again demonstrated the SE bars overlapping, confirming the difference between the means is not statistically significant.

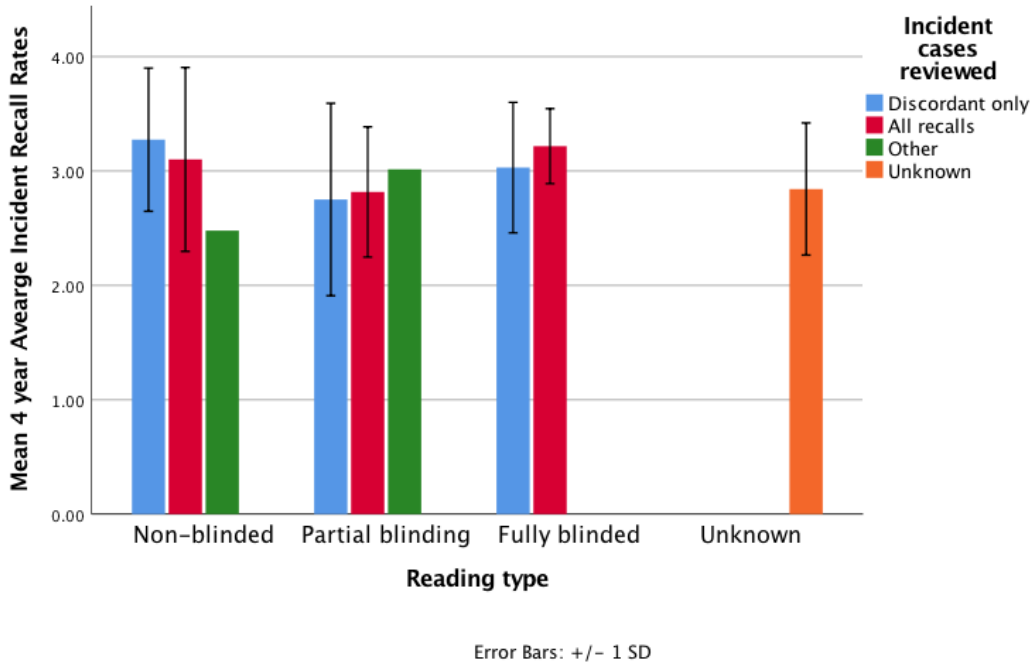


Figure 40. Clustered Bar Mean of 4-Year Average Incident Recall, Reading Type and Incident Cases Reviewed.

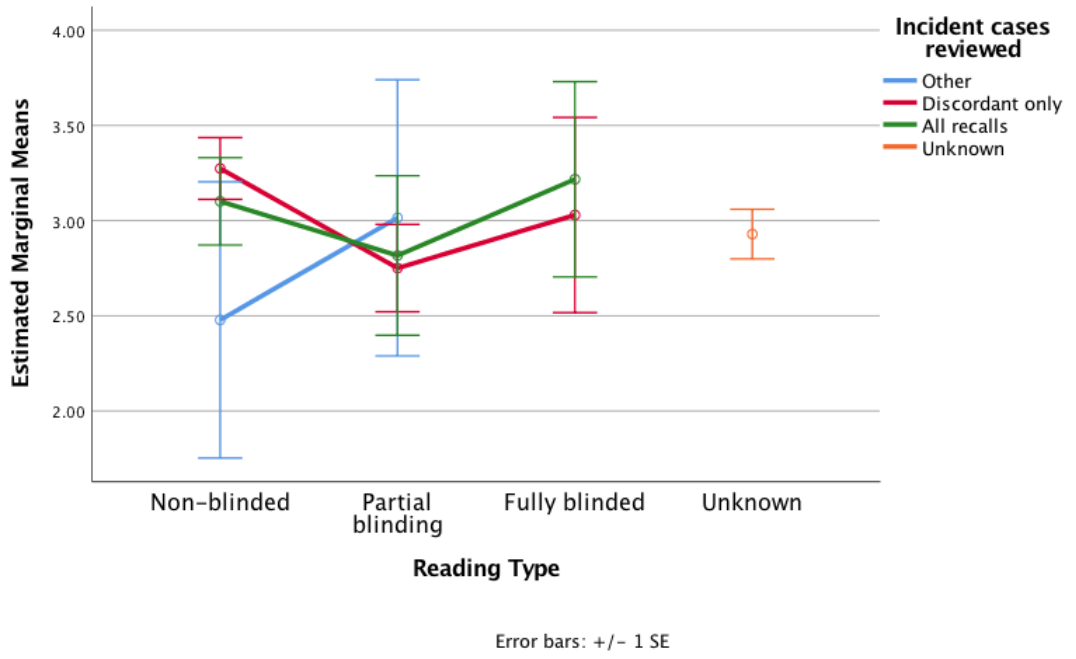


Figure 41. Estimated Marginal Means of 4-Year Average Incident Recall Rates

Analysis of the main effect was performed for reading type and indicated there was no statistically significant main effect, F (2, 71) = 0.084, p = .920, partial η^2 = .002 (outlier included) F (2, 70) = 0.102 p = .904, partial η^2 = .003 (outlier excluded). Analysis of the main effect for cases reviewed was also not statistically significant F (2, 71) = .103, p = .903, partial η^2 = .003 on prevalent recall rates (outlier included) F (2, 70) = .125, p = .883, partial η^2 = .004 (outlier excluded). The 'unknowns' were also removed from the incident data and the procedure repeated. The results still show no main effect for reading type F (2, 41) = 0.087, p = 0.917, partial η^2 = 0.04 or cases reviewed F (2, 41) = 0.107, p = 0.899, partial η^2 = 0.005.

7.10 ANOVA

7.10.1 Radiographer Third Reader Arbitrators and Four Year Average Overall Recall Rates

From the survey responses, some units indicated that Radiographers were currently undertaking single third reader arbitration. One-way ANOVA was conducted to determine if the overall recall rates for the same 4-year period (2013-2017) were different depending upon Radiographers or Radiologists/Breast Clinicians undertaking the task. Survey responses indicating Radiographer arbitration was undertaken in their unit were amalgamated to include Advanced Practitioners and Consultant Radiographers. Therefore, there were four classified groups: Yes, Radiographers undertake third reader arbitration (n = 11), no, Radiographers do not undertake third reader arbitration (performed by Radiologists or Breast Clinicians n=15), not applicable for units only using group consensus (n=23) and the unknowns (n=31).

With the outlier unit excluded, data was normally distributed for each group, as assessed by the Shapiro-Wilk test (p >.05). Homogeneity of variance was met (Levene's test p = .071). Data are presented as mean ± standard deviation. Overall recall rates increased from the unknowns (3.6 ± 0.8), to the N/A (3.7 ± 1.0), to units not utilising Radiographer third reader arbitrators (4.0 ± 0.7) to units utilising Radiographer arbitrators (4.0 ± 0.8) groups, but the differences between the reading types were not statistically significant, F (3, 75) = 1.022, p = 0.388 (outlier excluded) F (3, 76) = 0.519, p = 0.670 (outlier included) as demonstrated in Figure 42.

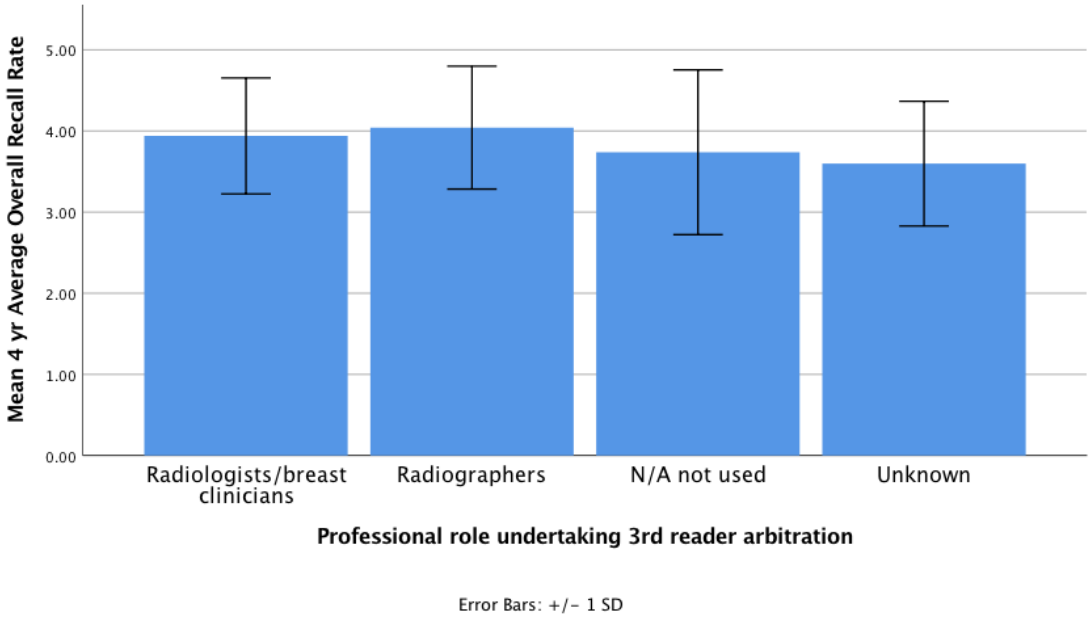


Figure 42. Histogram showing the Mean +/- SD of 4-Year Average Overall Recall Rate Relative to Units Utilising Radiographer Third Reader Arbitration

7.10.2 Radiographer Leading Consensus and Four Year Average Overall Recall Rates

The survey responses also confirmed that some units utilised Radiographers to lead/co-ordinate consensus group reviews. One-way ANOVA was conducted to determine if the overall recall rates were different depending upon Radiographers (Advanced Practitioners and Consultant Radiographers) leading consensus compared to other professional roles (Radiologists and Breast Clinicians). Some survey respondents selected that there was no leader of the consensus. Therefore, there were five classified groups: Yes, Radiographers lead consensus (n = 19), no, Radiographers do not lead consensus (n=11), no-lead (n=8), not applicable for units only using arbitration (n=11) and the unknowns (n=31). As before, with the outlier unit excluded data was normally distributed for each group, as assessed by Shapiro-Wilk test ($p > .05$). Homogeneity of variance was met (Levene's test $p = .343$). Data

are presented as mean ± standard deviation. Overall recall rates increased from the unknown group (3.6 ± 0.8), to the no lead group (3.6 ± 1.0), to units utilising Radiographers as leads of consensus (3.8 ± 0.7) to units not using consensus (4.0 ± 0.8) to units using only Radiologists or Breast Clinicians to lead (4.1 ± 1.1). However, the differences between the groups were not statistically significant, F (4, 74) = 0.962, p = 0.433 (outlier excluded) F (4, 75) = 0.554, p = 0.696 (outlier included) as demonstrated in Figure 43.

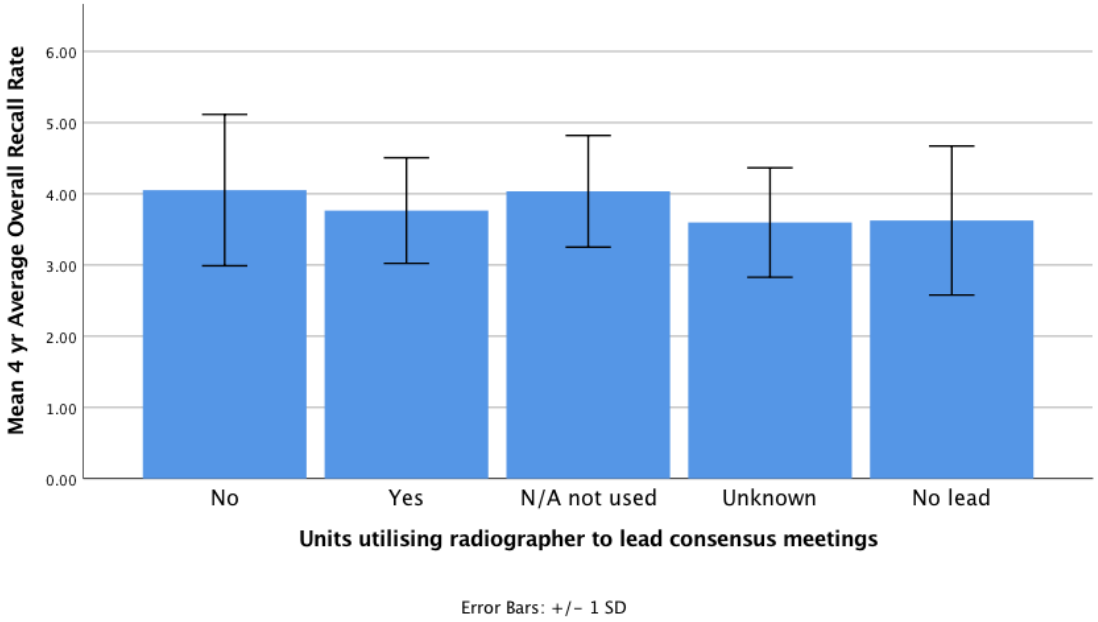


Figure 43. Histogram showing the Mean +/- SD of 4-Year Average Overall Recall Rate relative to Professionals Leading Consensus Meetings

7.10.3 Programme Size and Four Year Average Overall Recall Rates

Results from the Blanks et al. (2002) study implied that performance in smaller programmes was slightly inferior compared to medium and large programmes as measured by PPV of assessment and cancer detection rates. PPV is the likelihood of invasive cancer being present when recalled for assessment. As recall rates can be

significantly influenced by third reader arbitration or consensus, one-way ANOVA was primarily conducted (data 2013-2017) to determine if overall recall rates were different with the size of the program.

Programme size was classified into three groups: small (n = 20), medium (n=40), and large (n=20). The one same unit which (small size category) was an extreme outlier in the data as assessed by inspection of a boxplot. The analysis was therefore again conducted both with and without this unit included. Data were normally distributed for each group, as assessed by the Shapiro-Wilk test (p >.05) with the outlier excluded. Variances were homogeneous, as assessed by Levene's test for equality of variances (p = .448). Data are presented as mean ± standard deviation. Overall recall rates increased from the small programmes (3.6 ± 0.7), to the medium programmes (3.7 ± 0.9), to large programmes (4.0 ± 0.9), but the differences between the programme sizes were not statistically significant, F (2, 76) = 1.337, p = 0.269 (outlier excluded) F (2, 77) = 0.576, p = 0.564 (outlier included) as demonstrated in Figure 44.

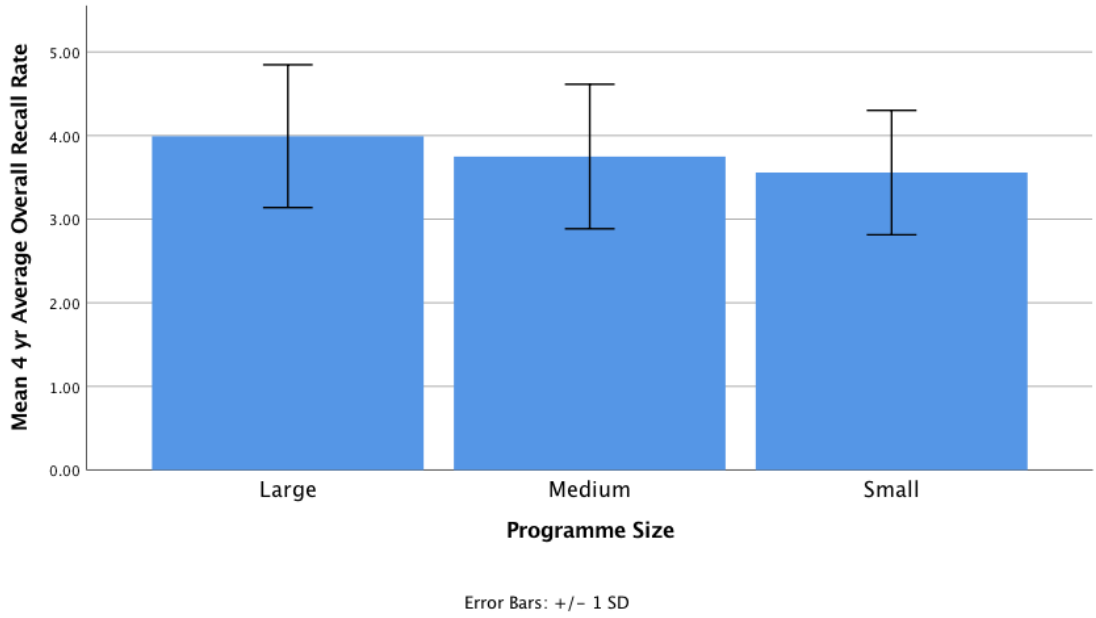


Figure 44. Histogram showing the Mean +/- SD of 4 Year Average Overall Recall Rate by Programme Size

7.10.4 Programme Size and Four Year Average Prevalent and Incident SDR

One-way ANOVA was also conducted (data 2013-2017) to determine if prevalent and incident SDR were different with the size of the program. With the same programme size classification analysis was conducted both with and without the extreme outlier unit included.

Prevalent data was normally distributed for each group, as assessed by the Shapiro-Wilk test (p >.05) with the outlier excluded. The assumption of homogeneity of variances was violated, as assessed by Levene's test for equality of variances (p = .001). Data are presented as mean ± standard deviation. Prevalent SDR rates increased from the small programmes (1.5 ± 0.3) to the large programmes (1.6 ±

0.2), to medium programmes (1.6 ± 0.2), but the differences between the programme sizes were not statistically significant, Welch F (2, 34.126) = 0.889, p = 0.420 (outlier excluded).

Incident data was normally distributed for each group, as assessed by the Shapiro-Wilk test (p >.05) with the outlier excluded. Variances were homogeneous as assessed by Levene's test for equality of variances (p = .662). Data are presented as mean ± standard deviation. Incident SDR rates increased from the small programmes (1.47 ± 0.1), to the large programmes (1.47 ± 0.1), to medium programmes (1.48 ± 0.1), but the differences between the programme sizes were not statistically significant, F (2, 76) = 0.135, p = 0.874 (outlier excluded).

7.11 ANCOVA

The one-way ANCOVA (analysis of covariance) is considered an extension of ANOVA and similarly can be used to ascertain if there are any significant differences between two (or more) independent groups on the dependent variable (Leppink 2018). Compared to the one-way ANOVA, the one-way ANCOVA allows statistical control for a third variable (often termed the confounding variable). This third variable that may confound results is called the covariate**.** Therefore, to determine whether small cancer detection rates and SDR (prevalent and incident screens) differed based on the arbitration strategy while controlling for the recall rate, one-way ANCOVA was conducted.

7.11.1 Four Year Average Prevalent Recall Rate, Four Year Average Prevalent <15mm CDR's and Arbitration Strategy

An ANCOVA was run to determine the effect of the arbitration strategy on prevalent small cancer detection rates after controlling for prevalent recall rates. Figures 45-48 demonstrate the range of the 4-year average prevalent recall rates (4.19-13.71) and the 4-year average prevalent small CDR (0.75-5.20) mapped to the individual arbitration strategy. The R-squared value is < 0.3 for all strategies, which is considered a none or very weak effect size.

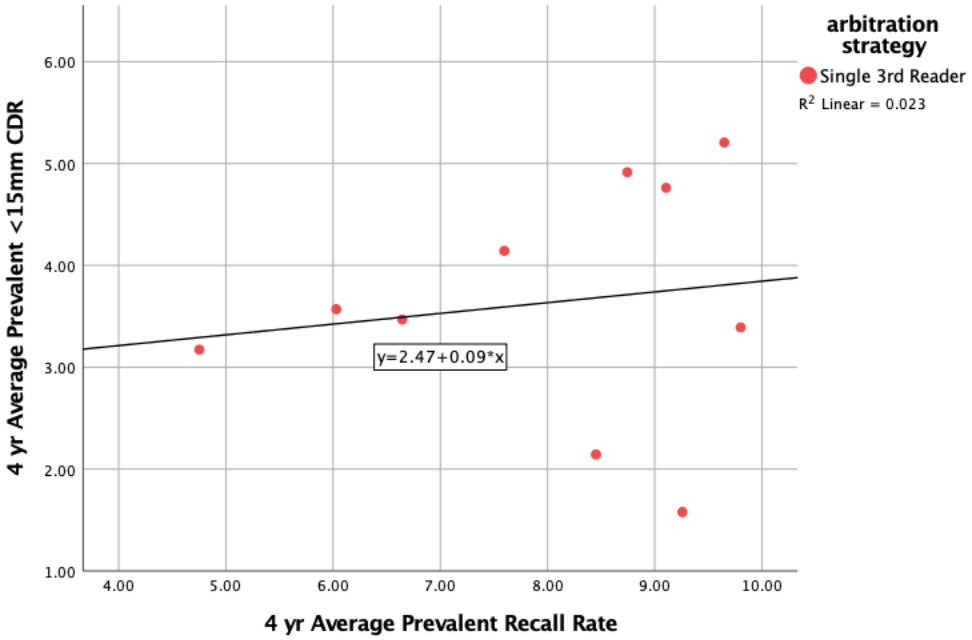


Figure 45. Grouped Scatter plot of Prevalent <15mm CDR, Prevalent Recall Rates and 3rd Reader Arbitrator.

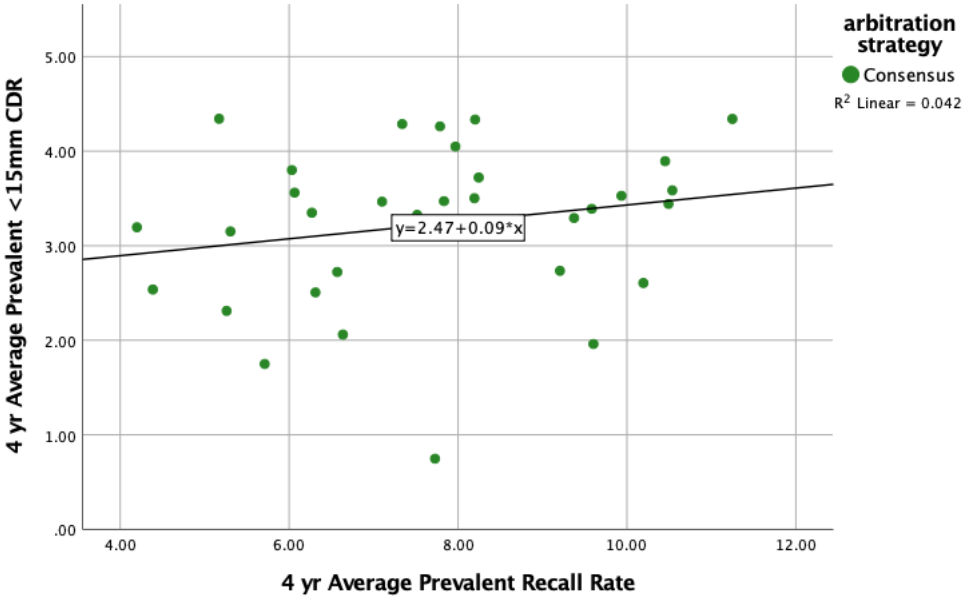


Figure 46. Grouped Scatter plot of Prevalent <15mm CDR, Prevalent Recall Rates and Consensus.

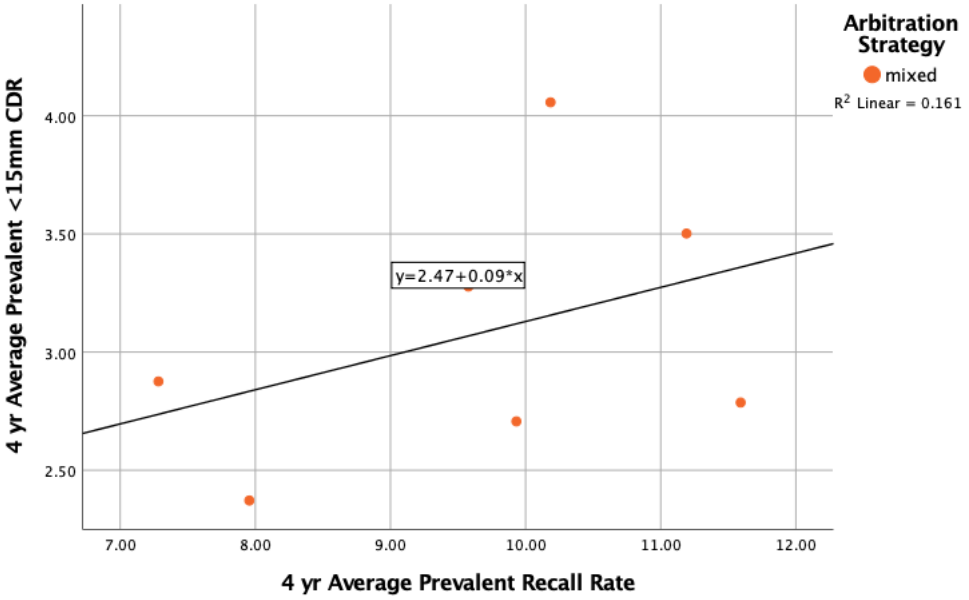


Figure 47. Grouped Scatter plot of Prevalent <15mm CDR, Prevalent Recall Rates and Mixed Strategy.

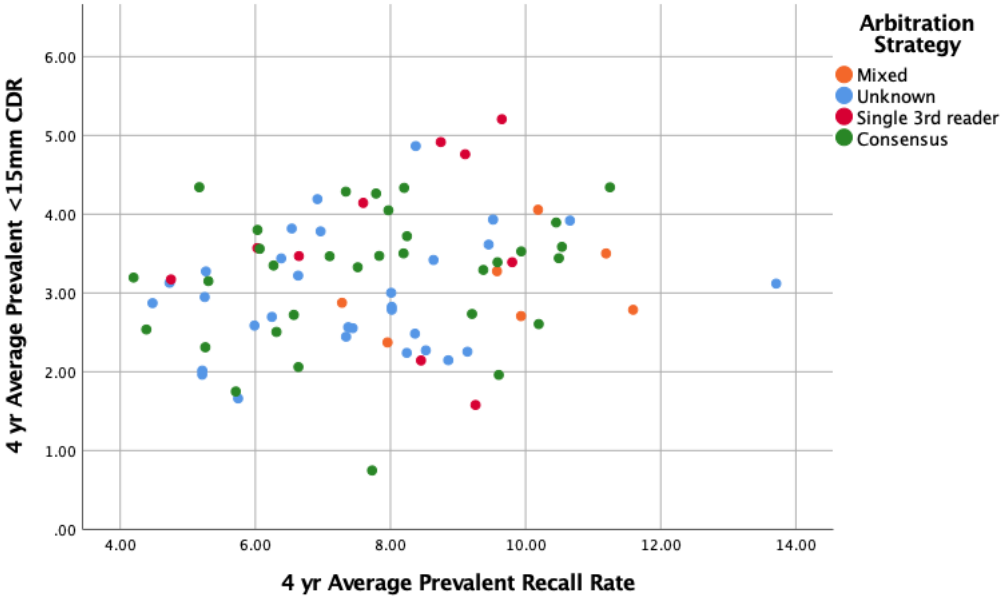


Figure 48. Grouped Scatter plot of Prevalent <15mm CDR, Prevalent Recall Rates and All Arbitration Strategies (including the unknowns).

There was homogeneity of regression slopes as the interaction term was not statistically significant, F (3, 72) = 0.026, p = 0.994. Standardised residuals for the interventions were normally distributed as assessed by the Shapiro-Wilk's test (p > .05). There was homoscedasticity, as assessed by visual inspection of the standardised residuals plotted against the predicted values. Variances were homogeneous as assessed by Levene's test of homogeneity of variance (p = 0.291). There was one outlier in the data, as assessed by cases with standardised residuals greater than ±3 standard deviations. Data are adjusted mean ± standard error. Small CDR was greater in the third reader arbitrator group (3.62 ± 0.26) compared to the consensus group (3.24 ± 0.15) the unknown group (3.00 ± 0.15) and the mixed strategy group (2.91 ± 0.32), respectively. After adjustment for recall rates, there was not a statistically significant difference in small cancer detection rates between

the strategies, F (3, 75) = 1.72, p = 0.17, partial η^2 = .064. A one-way ANCOVA was rerun without the outlier included in the analysis. In conclusion, both results show no statistically significant difference F (3, 74) = 2.21, p = 0.09, partial η^2 = 0.082.

7.11.2 Four-year Average Prevalent Recall Rate, 4-Year Average Prevalent SDR and Arbitration Strategy

Figures 49-52 demonstrate the range of the 4-year average prevalent recall rates (4.19-13.71) and the 4-year average prevalent SDR (0.94-2.15) mapped to the individual arbitration strategy. Again, the R-squared value is < 0.3 for all strategies, which is considered a none or very weak effect size.

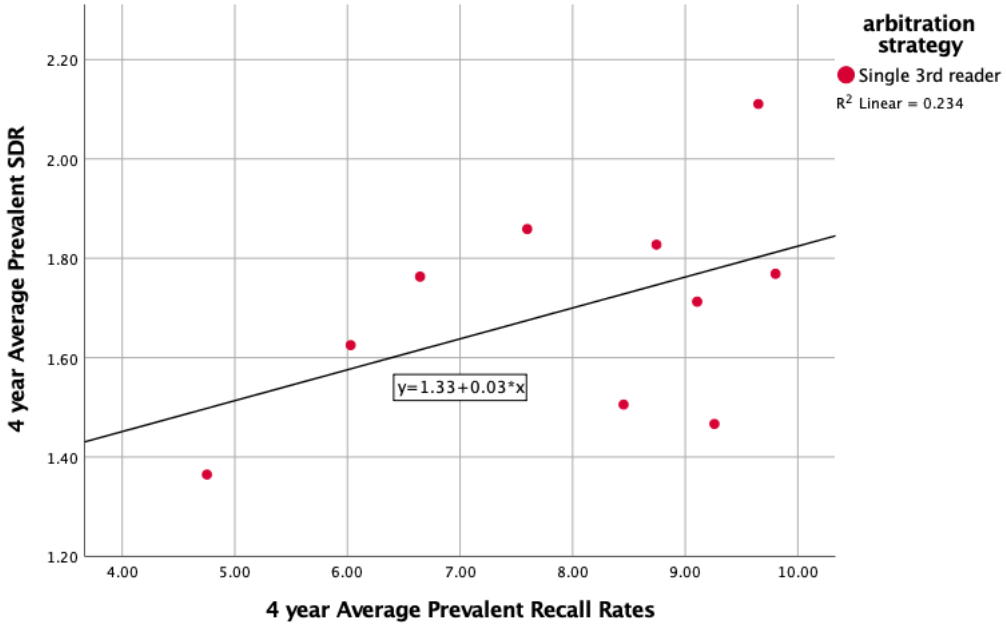


Figure 49. Grouped Scatter plot of Prevalent SDR, Prevalent Recall Rates and 3rd Reader Arbitrator

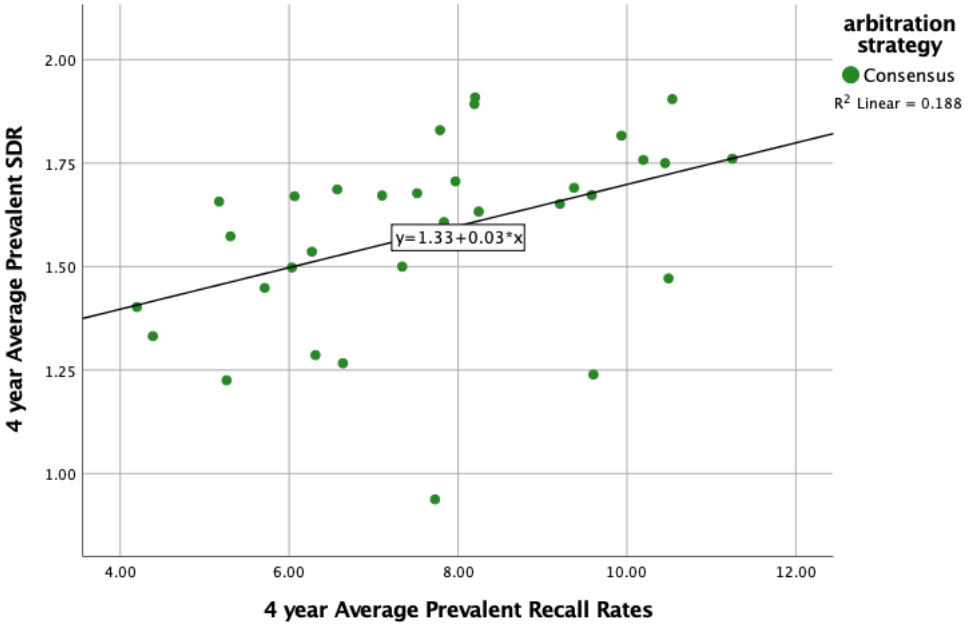


Figure 50. Grouped Scatter plot of Prevalent SDR, Prevalent Recall Rates and Consensus

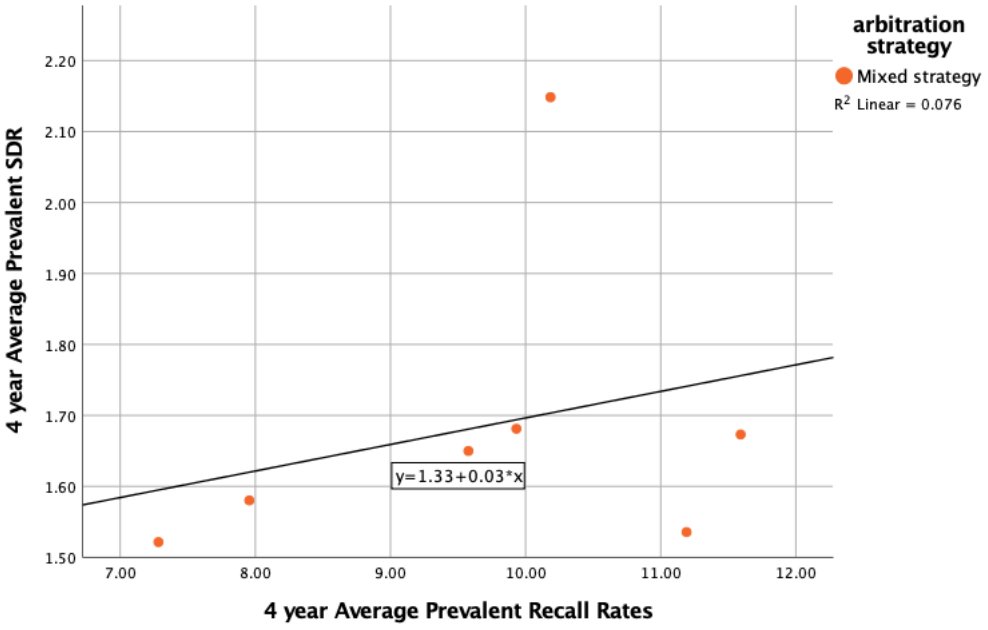


Figure 51. Grouped Scatter plot of Prevalent SDR, Prevalent Recall Rates and Mixed Strategy

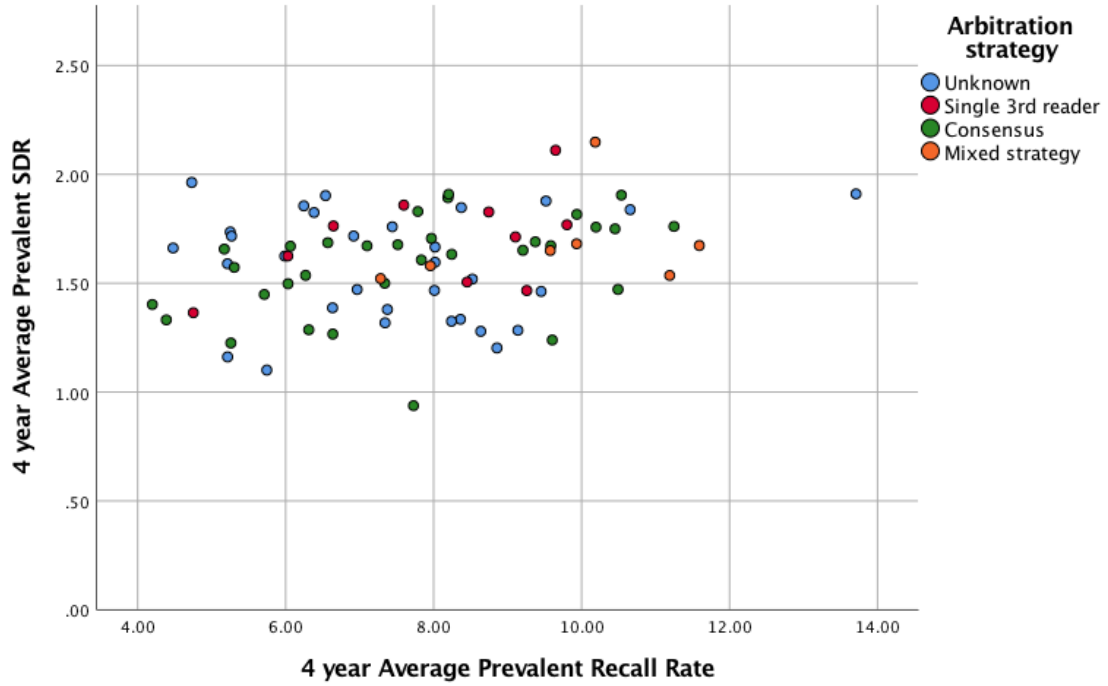


Figure 52. Grouped Scatter of Prevalent SDR, Prevalent Recall Rates and All Arbitration Strategies (including the unknowns).

Analysis of prevalent recall rates and prevalent SDR demonstrated there was homogeneity of regression slopes as the interaction term was not statistically significant, F (3, 72) = 0.904, p = 0.443. Standardised residuals for the interventions were not normally distributed, as assessed by the Shapiro-Wilk's test (p < .05) for the consensus and mixed strategy groups. However, one-way ANCOVA is reported to be reasonably robust to deviations from normality with heterogeneity having a more significant effect on F-test robustness than non-normality (Blanca et al. 2017) and therefore the test was still considered appropriate. There was homoscedasticity, as assessed by visual inspection of the standardised residuals plotted against the predicted values. There was homogeneity of variances, as assessed by Levene's test of homogeneity of variance (p = .090). There were no

outliers in the data, as assessed by no cases with standardised residuals greater than ±3 standard deviations.

Data are adjusted mean ± standard error. Prevalent SDR was higher in the third reader arbitrator group (1.69 ± 0.7) compared to the mixed group (1.63 ± 0.09) the consensus group (1.59 ± 0.04) and the unknown strategy group (1.59 ± 0.04), respectively. After adjustment for recall rates, there was no statistically significant difference in prevalent SDR between the strategies, F (3, 75) = 0.667, p = 0.58, partial η^2 = .026

7.11.3 Four Year Average Incident Recall Rate, Four Year Average Incident <15mm CDR's and Arbitration Strategy

Data for the same period was run in ANCOVA to determine the effect of the arbitration strategy on incident small cancer detection rates after controlling for recall rates. Figure 53 demonstrates the range of the 4-year average incident recall rates (1.58-5.54) and the 4-year average incident small CDR (2.32-4.43) mapped to the arbitration strategy.

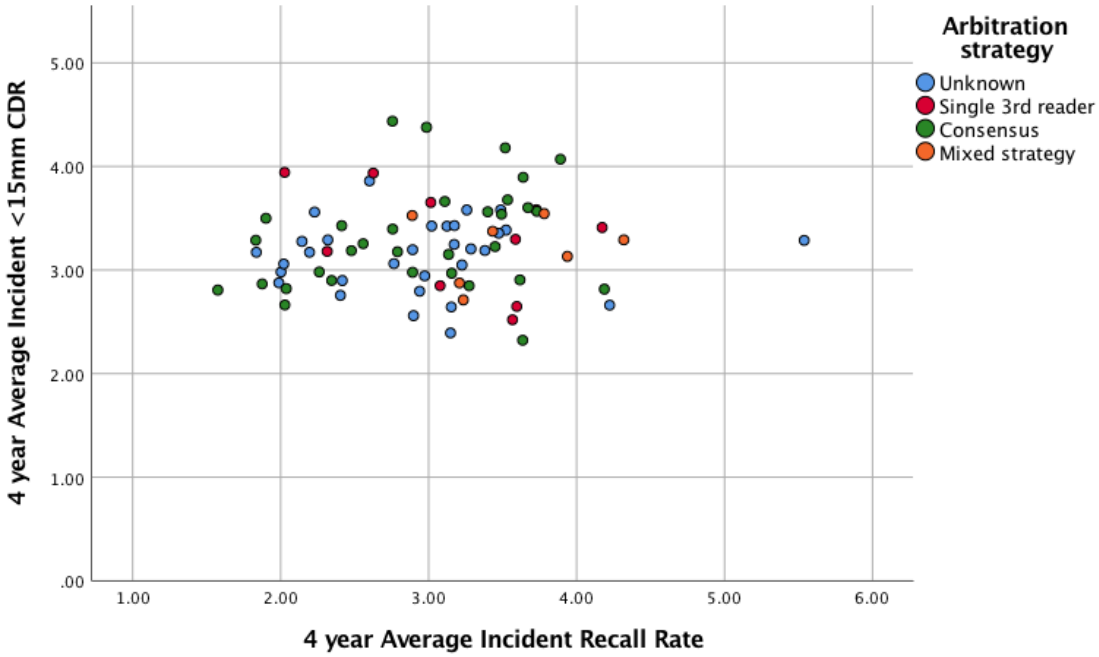


Figure 53. Grouped Scatter of Incident <15mm CDR, Incident Recall Rates and Arbitration Strategy

There was homogeneity of regression slopes as the interaction term was not statistically significant, F (3, 72) = 1.66, p =0.183. Standardised residuals for the interventions were normally distributed, as assessed by Shapiro-Wilk's test (p > .05). There was homoscedasticity, as assessed by visual inspection of the standardised residuals plotted against the predicted values. There was homogeneity of variances, as assessed by Levene's test of homogeneity of variance (p = .122). There were no outliers in the data, as assessed by no cases with standardised residuals greater than ±3 standard deviations. Data are adjusted mean ± standard error. Incident <15mm CDR was higher in the consensus group (3.32 ± 0.08) compared to the third reader arbitrator group (3.29 ± 0.14) the mixed group (3.18 ± 0.17) and the unknown strategy group (3.14 ± 0.08), respectively. After adjustment for recall rates, there

was no statistically significant difference in small cancer detection rates between the strategies, F (3, 75) = 0.957, p = 0.418, partial η^2 = .037.

7.11.4 Four Year Average Incident Recall Rate, Four Year Average Incident SDR and Arbitration Strategy

An ANCOVA was initially run to determine the effect of the arbitration strategy on incident SDR after controlling for recall rates. Figure 54 demonstrates the range of the 4-year average incident recall rates (1.58-5.54) and the 4-year average incident SDR (1.21-1.79) mapped to the arbitration strategy. One of the data assumption checks of ANCOVA is that there is no interaction between the covariate (incident recall rate) and the independent variable (arbitration strategy) (i.e. the regression lines must be parallel having the same slope). The results demonstrated there was a statistically significant interaction ($p < .05$), and therefore the assumption of homogeneity of regression slopes was violated, and ANCOVA analysis was not appropriate (Johnson and Edu 2016).

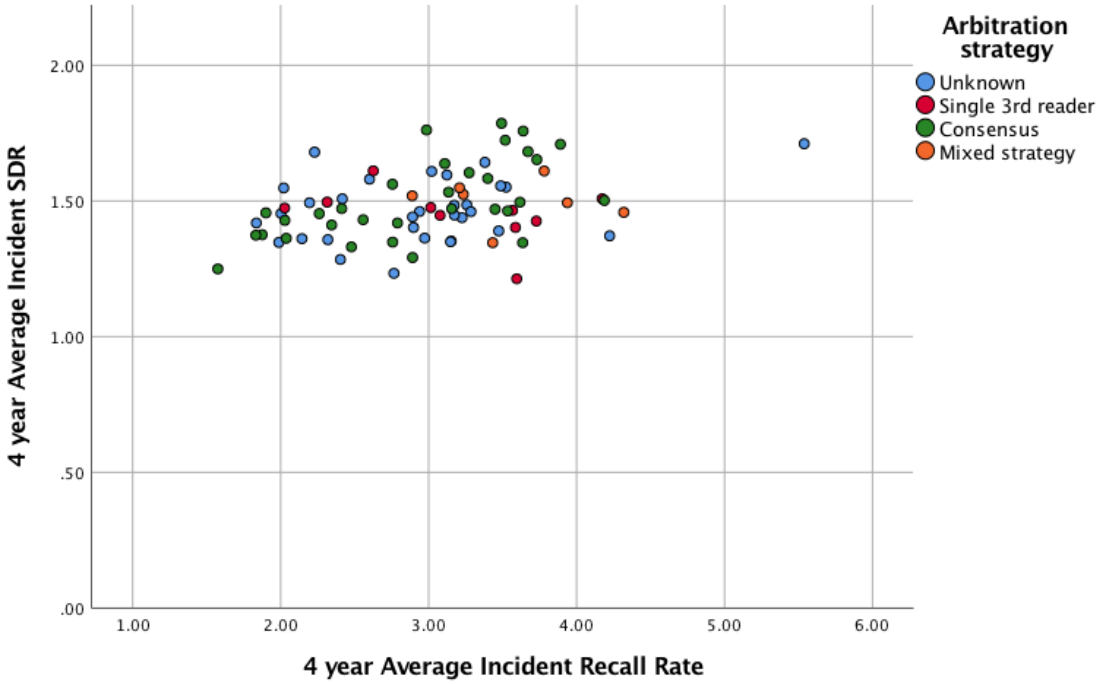


Figure 54. Grouped Scatter of Incident SDR, Incident Recall Rates and Arbitration Strategy

A multiple regression was, therefore, run to determine how much of the variation in incident SDR can be explained by the incident recall rates and incident arbitration strategy as a whole, but also the relative contribution of each of the independent variables in explaining the variance. There was linearity as assessed by partial regression plots and a plot of studentised residuals against the predicted values.

Independence of observations in multiple regression is designed to

> '*test for 1st-order autocorrelation, which means that adjacent observations are not independent*' (Schneider, Hommel, and Blettner 2010).

In SPSS, this can be checked using the Durbin-Watson statistic. This statistic can range from 0 to 4, with a value of approximately two signifying that there is no correlation between residuals. Residuals were independent as assessed by a Durbin-Watson statistic of 1.55. There was homoscedasticity, as assessed by visual inspection of a plot of studentised residuals versus unstandardised predicted values.

Multicollinearity occurs when the independent variables are highly correlated, with a change in one variable, causing a change in another variable. This is problematic and may cause difficulties in interpretation of the results (biased estimation and the statistical power of the regression model may be reduced) (Yoo et al. 2014). There was no evidence of multicollinearity, as assessed by tolerance values greater than 0.1. There were no outliers (studentised deleted residuals higher than ±3 standard deviations).

The data was also assessed to determine whether any cases exhibited high leverage which can influence the regression analysis. Huber and Ronchetti (2009) consider leverage values '*less than 0.2 as safe, 0.2 to less than 0.5 as risky, and values of 0.5 and above as dangerous'*. In this data set, two units had a 0.2 value. However, Cook's Distance measure, which provides an indication of influence on a data point was 0.14. A Cook's Distance greater than 0.5 requires further investigation as it may be influential (Cook and Weisberg 1982). To run inferential statistics (i.e., determine statistical significance), the errors in prediction (residuals) must be normally distributed. The assumption of normality was met, as assessed by a P-P Plot.

The multiple regression model statistically significantly predicted incident SDR *F* (4, 75) = 4.57, *p*= 0.002, adj. R^2 = .153. All four variables added statistically significantly to the prediction, $p < .05$. Regression coefficients and standard errors can be found in Table 40. This demonstrates that for a 1% increase in incident recall rates there was an increase in incident SDR which was greatest for the consensus strategy (0.08) and least for the third reader arbitrator (0.05).

Table 40. Summary of Multiple Regression Analysis

Model	Unstandardised Coefficients		Standardised Coefficients
	B	Std. Error	Beta
Intercept	1.278	.057	
Unknown strategy	.062	.020	.753
3rd reader arbitrator	.050	.021	.436
Consensus	.080	.020	.965
Mixed strategy	.061	.020	.498

7.12 Correlation of Resources for Consensus Group Review, Third Reader Arbitrator (Time, number of cases, number of staff and number of sessions) and Four Year Average Overall Recall Rates

Results from the survey indicated that more units undertake consensus group review compared to a single third reader arbitrator. However, free-text comments highlighted that consensus meetings often included one or both of the original readers who rarely changed their opinion and subsequently, very few women were returned to routine recall. Thematic analysis of free-text comments also emphasised that there are limited staffing resources, that group consensus is time-consuming,

resource-intensive and may be logistically challenging. Hence, the data was analysed to determine if the time and resources invested in group consensus correlated to a reduction in overall recall rates.

A Pearson's correlation was initially run to assess the relationship between total staff minutes (number of staff per meeting multiplied by the time per week) and 4-year average overall recall rates. Not all variables were normally distributed, as assessed by Shapiro-Wilk's test ($p < .05$) and therefore, the non-parametric Spearman's rank correlation was also undertaken.

The results demonstrate there was no statistically significant correlation between total staff minutes and overall recall rates, Spearman $r_s(47) = .239$, $p = .098$. This suggests that increasing time and number of staff spent in reviewing arbitration cases does not impact on overall recall rates.

A Pearson's correlation was also run to assess the relationship between total staff minutes (number of staff per meeting multiplied by the time per week) and 4-year average prevalent and incident SDR. Again, not all variables were normally distributed, as assessed by Shapiro-Wilk's test ($p < .05$) and therefore, Spearman's rank correlation was performed. There was no statistically significant correlation between total staff minutes and 4-year average prevalent SDR, Spearman $r_s(47) = -0.150$, $p = 0.305$. This was also the case for incident SDR Spearman $r_s(47) = -0.072$, $p = 0.624$. This again suggests that increased resources (time and number of staff) spent reviewing cases, does not impact on overall SDR rates.

7.13 Conclusion

In this study, there are variations in the performance parameters reviewed at the unit level. In particular, recall rates, the reasons for which are unclear. Recalling a higher proportion of normal women for assessment adds additional pressures to services already facing staffing shortages.

No statistically significant correlation was found between the 4-year average recall rates and small (<15mm) cancer detection rates (prevalent and incident). However, there was a statistically significant, weak/moderate positive correlation between the prevalent/incident recall rates respectively and SDR. The peak incident SDR in the data occurred with a recall rate of 3.781%.

In this study, there was no difference in mean recall rates between units for the cases reviewed; the arbitration strategy; the reading type; professional role undertaking the third reader arbitration/leading consensus or programme size. There were no statistically significant differences for the four-year average prevalent and incident SDR between programme sizes. In conclusion, overall, there was no statistically significant difference between the arbitration strategies for small cancer detection rates (prevalent and incident) or SDR (prevalent and incident).

Increasing time and number of staff spent in reviewing arbitration cases, did not impact on overall recall rates or overall SDR. In units with a diminishing workforce reverting to a third reader may become more common as consensus group review was considered time-consuming by some respondents in the survey. However, for a 1% increase in incident recall rates, there was an increase in incident SDR which was

greatest for the consensus strategy (0.08) and least for the third reader arbitrator strategy (0.05). The slightly higher increase obtained with consensus may reflect that a group review makes the most of skill mix and precludes the 'blind spots' individual to each reader.

There are several limitations in this study which could affect the results. The individual round length for each unit is unknown and hence any slippage in round length (women are screened more than three years since their last screen) could affect the units SDR. The KC62 data publicly available does not provide data on non-invasive or micro-invasive disease for individual units. It was therefore not possible to ascertain if higher recall rates were associated with higher DCIS rates.

Reading practices, arbitration strategies, and Radiographer arbitration data are only available for units that responded to the surveys and therefore, analysis of this data for all 80 units may show different and statistically significant results. Some units introduced Radiographer arbitration within 2016-2017 and therefore a review of national data over an extended period would need to be analysed to ascertain if there is any impact on overall recall rates.

The next chapter builds upon the information obtained from the surveys on current reporting/arbitration practices and the limitations of the current decision-making strategies. Semi-structured telephone interviews were undertaken to critique the PHE arbitration guidance and to explore staff perceptions on the barriers and facilitators to implementation. Also, the interviews were used to explore alternative models of service delivery and the future use of AI in breast screening.

Chapter 8. Semi-Structured Telephone Interviews

8.1 Introduction

The literature review has demonstrated the complexity of human decision-making and the dynamics associated with group decision-making. The survey results in Chapter 6 also illustrate national variance in all elements of reporting and arbitration practices, with a potential impact on the second reader and recall rates. However, although the KC62 data analysis demonstrated variance in recall rates, this could not be attributed to a particular reporting or arbitration strategy.

This chapter discusses the qualitative research undertaken, using semi-structured telephone interviews, offering a rationale for the sample selected from the pre-determined sampling frame, while critically appraising the data collection method. The interviews conducted with the reporting staff aimed to explore further the rationale for currently observed variances, and views on the subsequent effects on decision-making, recall rates and ultimately CDR and how it might be improved. Participants' views on alternative models of service delivery, and the information technology required to support reporting and arbitration practices in the future were also sought.

8.1.1 Rationale

Qualitative research methods are able to focus on processes and can frequently highlight a sense of change (Bryman 2015). Both are relevant to this study following the publication of the PHE arbitration guidance. Individual interviews can provide an opportunity for in-depth exploration of an individual's personal views of the topic being studied and enable detailed subject coverage (Coolican 2004). Ritchie and

Lewis (2003) state that individual interviews are appropriate for research that requires an understanding of entrenched or complex systems. This approach was therefore judged appropriate to this study, as Chapter 4 identified that in an NHS organisation, systems of work and behaviour can often be embedded, with organisational culture being the main factor that inhibits change (Catchpole 2013).

8.1.2 Research Objectives of Semi-Structured Interviews

Chapter 6 established that non-blinding reading and arbitration is predominantly undertaken (in the respondent units) with potential repercussions on the decision-making of the second reader and third reader arbitrator (or group). In-depth interviews were therefore conducted to explore the survey data further in order to: investigate the limitations of the current reporting/arbitration practices; study staff opinions on the content of the PHE arbitration guidance; understand the motivations to implement or to negate Radiographer arbitration; explore the impacts and outcomes of unit variance and the future use of AI in breast screening.

8.2 Methods

8.2.1 Telephone Interviews

Some social science literature concludes that face-to-face interviews are superior for constructing narrative data, with other modes (e.g. telephone) considered inferior (Holt 2010, Irvine, Drew, and Sainsbury 2013, Kazmer and Xie 2008). Key challenges described relate to difficulties in establishing a rapport, the inability to react to visual cues, and observation of the individual in the work environment; possibly diminishing the quality of empirical data collected. Conversely, Stephens (2007) and Cachia and Millward (2011) report telephone interviews are constructive and a valid

methodological tool. Specifically, telephone interviews offer a potential increase in participants' availability, allow targeting of a wide geographical area, reduce researcher time and travel costs, and provide greater flexibility for scheduling compared to face-to-face interviews. In this study, a representative sample of participants throughout England was required, so telephone interviews offered a practical solution for interviewing busy staff with clinical and professional priorities. Also, several authors (Lechuga 2012, Cachia and Millward 2011, and Stephens 2007) emphasise other benefits of telephone interviews which include increased privacy, reduced distractions (for interviewees), less self-consciousness during note-taking (for interviewers) and perceived anonymity. Furthermore, the absence of non-visual cues can be considered advantageous as the conversation needs to be clearly articulated by both individuals and a richer text results from which to commence data analysis (Stephens 2007). The power dynamics that may exist in a face-to-face interview between the researcher and the interviewee may also be negated (Muntanyola Saura and Romero Balsas 2014, Holt 2010). In this study, staff of varying clinical positions were interviewed (senior to the researcher, peers and staff of lower clinical grades) but all possessed a shared body of knowledge through their related professional roles. A telephone interview can also be considered less intrusive, offering greater control to the participants allowing termination of the interview (Muntanyola Saura and Romero Balsas 2014, and Holt 2010). Therefore, given the nature of the participant's job role, and likely availability, telephone interviews were employed as the most pragmatic option. Participants were not offered a face-to-face interview given the geographical distance to be travelled.

8.2.2 Semi-Structured Interviews

Qualitative interviews can be structured, semi-structured, or unstructured, depending on the requirements (Saunders, Lewis, and Thornhill 2009). Structured interviews include a firmly defined set of questions for each interview, with the interviewer controlling the conversation to a high degree. In contrast, unstructured interviews contain a limited number of open-ended questions related to the topic being examined (Saunders, Lewis, and Thornhill 2009), providing the interviewee considerable freedom to express their views. Semi-structured interviews lie between the two extremes and are designed with a set of central questions that can be adapted appropriately to the situation (Saunders, Lewis, and Thornhill 2009). Central questions are followed up by prompts allowing the researcher flexibility to gain rich descriptive data of the participant's personal experience while limiting the discussion to pertinent issues (Coolican 2004). Use of semi-structured interviews was deemed appropriate for this study as the survey findings had demonstrated different reporting and arbitration practices, with an apparent lack of evidence to support some of the historical, cultural practices. Participant views were therefore required to explore specific variations and their views on the processes used, PHE guidance, and difficulties in defining and monitoring quantitative guidance for arbitration.

8.2.3 Rigour of Qualitative Research

For quantitative research, there are defined methods for ascertaining the rigour and quality of a study (Murphy and Yielder 2010). Noble and Smith (2015) confirm that measures used to authenticate the validity and reliability of quantitative research are not relevant to qualitative studies. Validity refers to research measuring what it

actually aims to measure (Roberts, Priest, and Traynor 2006). Reliability refers to how well a test or tool (with no variation in other factors) produces consistent and dependable results in different circumstances (Murphy and Yielder 2010). In quantitative research, variables are more easily controlled. In qualitative research, the researcher may be a variable, and therefore control is more challenging. Murphy and Yielder (2010) identify that the fundamental principle of all the above is reflexivity. Reflexivity is the ability of a researcher to understand their position in a study and their relationship with participants (Santiago-Delefosse et al. 2016). Examples of the strategies used to ensure rigour and reflexivity in this study are shown in Table 20 Chapter 5.

8.2.4 Sample

Unlike the quantitative study, qualitative research adopts a different paradigm in which statistical representation and scale are less consequential (Mason 2017, and Patton 2015). Instead, the population from which the sample is drawn, the ability to denote relevant characteristics, and the quality of the information collated demonstrate the exactitude and rigour of a sample (Kelly 2012).

8.2.5 Stratified Purposive Sampling

Ritchie and Lewis (2003) describe different approaches (Table 41) that can be utilised in purposive sampling dependent upon the study aims. In purposive sampling, participants are selected based on specific criteria (Mason 2017). In this study, stratified purposive sampling was used to cover the views and experiences of different professional roles (Directors, Radiologists, Breast Clinicians, Consultant Radiographers and Advanced Practitioners) and allow a comparison of these

subgroups. All interviewees had to have completed the survey and indicated in their survey response that they were willing to be followed up.

Table 41. Varying Approaches to Purposive Sampling (Taken from Ritchie and Lewis 2003)

• Homogeneous samples
Selected to give a comprehensive representation of a specific phenomenon – e.g. individuals who possess the same characteristics. Enables detailed exploration in a particular context.
• Heterogeneous samples
An intentional stratagem to include phenomena with a wide variation. Enables the greatest variation in sampling. Aim to identify predominant themes across a variety of people.
• Deviant sampling (extreme cases)
Cases selected as they are uncommon or remarkable. Theoretically informative, learning about the phenomena is enhanced by studying exceptions/extremes.
• Intensity sampling
Cases selected which compellingly represent the phenomena of interest.
• Typical case sampling
Cases which portray 'normality' are selected to provide detailed depictions (e.g. interviewees may be selected from survey responses).
• Stratified purposive sampling *(utilised in this study)*
Groups which exhibit a variation of the phenomena but each of which is relatively homogeneous allowing comparison of subgroups.
• Critical case sampling
Cases selected based on the logic that they validate the phenomenon and are crucial to the interpretation proposed by the research.

As stratified purposive sampling entails deliberate selection by the researcher, it is, therefore, essential to demonstrate clear objectivity and avoid biased selections (Ritchie, Jane and Lewis 2003). This was achieved by the following pre-determined sampling frame.

8.2.6 Justification of the Sample

A pre-determined sampling frame was constructed based on KC62 unit performance data (2015-2016 latest published at the time) and individual characteristics (professional role, arbitration strategy). Two criteria were selected for unit performance; overall recall rates, and incident small (<15mm) cancer detection rate. The overall recall rate was deemed an appropriate measure of unit performance as this can be significantly affected by the process of arbitration. Incident small cancer detection rate was used as the second parameter because detection of small invasive cancer is expected to reduce breast cancer mortality (Tabàr et al. 1992) compared to in situ or large invasive cancers. Larger cancers are less likely to be missed by either reader, and discordant cases requiring arbitration review are more likely to be small cancers or subtle mammographic abnormalities. Also, there is less likely to be statistical variation in incident cancers as they should mainly be cancers that have developed in the three years since the prevalent screen (Duffy and Gabe 2005). Appendix 9 demonstrates the 4-year data for these criteria. The purpose of this sampling method was to ensure that all professional roles were interviewed from a range of units, i.e. high recall rates/high small CDR, high recall rates/low small CDR etc. In each primary sampling cell, the fundamental aim was to ensure there was diversity in professional roles and the strategies used to manage discordant cases.

In order to do this, overall recall rates were ordered and evenly divided into three groupings of low recall rates 1.99 -3.54 (27 units), medium recall rates 3.61-4.10 (26 units) and high recall rates 4.13 -7.04 (27 units). The same principle was applied to

incident small (<15mm) cancer detection rates, with low CDR categorised as 1.99 - 3.13, medium 3.15-3.66 and high 3.67-5.05. This data was imported into a 3 x 3 framework (Table 42) which automatically produced a spread of small, medium and large units across the nine cells. Although it is acknowledged that, for six units, the 2015-2016 CDR was based on small numbers, averaging the data over four years did not affect cell placement. The pre-determined sampling frame also ensured that particular strategies had an equal opportunity to be explored. A review of the sampling frame post-survey completion (Table 42) identified that in some specific cells, there were limited responses either by professional role or by arbitration strategy.

Table 42. 3 x 3 Sampling Framework. Size, Strategy and Professional Role of Respondent Units, Categorised by Recall Rates and Small CDR.

Low Recall Rate 1.99-3.54	Medium Recall Rate 3.61-4.10	High Recall Rate 4.13-7.04	
8 units - 1 small, 5 medium, 2 large. **Response from 4 units (50%)** 3 consensus and 1 mixed 2 consented to interview 1 Locum Radiologist (mixed) **Interviewed** 1 Locum Radiologist (mixed) (Large unit) 1 Consultant Radiographer(consensus) (Large unit)	**10 units -** 4 small, 3 medium, 3 large **Response from 3 units (30%)** 3 consensus 3 consented to interview 2 Directors (Consensus) 1 Consultant Radiographer(Consensus) **Interviewed** 1 Director (consensus) (medium unit) 1 Consultant Radiographer(consensus) (medium unit)	**9 units –**1 small, 4 medium, 4 large **Response from 6 units (67%)** 4 consensus and 2 arbitration Multiple roles consented to an interview but no arbitration **Interviewed** 1 Director (consensus) (Large unit) 1 Lead Radiographer (consensus) (small unit)	**Low <15mm CDR 1.99-3.13**
11 units - 3 small, 6 medium, 2 large **Response from 7 units (64%)** 7 consensus Multiple roles consented to an interview but no arbitration **Interviewed** 1 Director (consensus) (medium unit) 1 Advanced Practitioner (consensus)(small unit)	**10 units -** 0 small, 8 medium, 2 large **Response from 6 units (60%)** 2 consensus 2 mixed 2 arbitration Multiple roles consented to interview **Interviewed** 1 Director (arbitration) (large unit) 1 Advanced Practitioner (consensus)(medium unit)	**5 units -** 2 small, 2 medium, 1 large **Response from 4 units (80%)** 2 consensus 1 mixed 1 arbitration Multiple roles consented to interview **Interviewed** 1 Radiologist (consensus) (medium unit) 1 Consultant Radiographer (arbitration) (large unit)	**Medium <15mm CDR 3.15-3.663**
8 - units 5 small, 2 medium, 1 large **Response from 4 units (50%)** 2 consensus 2 arbitration Multiple roles consented to interview **Interviewed** 1 Director (arbitration) (medium unit) 1 Consultant Radiographer(arbitration) (medium unit)	**6 units -** 1 small, 2 medium, 3 large **Response from 4 units (67%)** 3 consensus 1 arbitration Multiple roles consented to interview **Interviewed** 1 Director (consensus) (large unit) 1 Breast Clinician (arbitration) (large unit)	**13 units-**3 small, 8 medium, 2 large **Response from 11 units (85%)** 6 consensus 3 mixed 2 arbitration Multiple roles consented to interview **Interviewed** 1 Director (consensus) (medium unit) 1 Radiologist (arbitration) (small unit)	**High <15mm CDR 3.67-5.05**

Predominantly, survey respondents reported group consensus (32 units) rather than a single third person arbitrator (10 units) or mixed strategy responses (7 units). Thus, prioritisation was given to ensuring interviews with different professionals (Director, Radiologist, Breast Clinician, Consultant Radiographers and Advanced Practitioners) followed by the process used (3rd reader arbitration or consensus). Since no Director had responded from a unit with a low recall rate/ low <15mm CDR category, this limited sampling from this category. The final interview sample (n=18) included 7 Directors, 2 Radiologists, one locum Radiologist (previously Director of another unit), 1 Breast Clinician (previously Director), 4 Consultant Radiographers and 3 Advanced Practitioner's/lead Radiographer. This sample covered eleven units which use group consensus, six using a single third reader arbitrator and one mixed strategy. In terms of unit size, there were seven large, eight medium and three small size units.

In addition to the stratified interview sample, three semi-structured telephone interviews were undertaken to gather further data on:

1. The historical practice of reporting and arbitration strategies on the National Breast Screening System (NBSS). This interview was undertaken with the previous Director of the NHS Cancer Screening Programmes (2015), and current Principal Investigator of the breast screening Age Extension trial whom it was considered would be able to provide a rationale for the evolution of practices.

2. The variance of third reader arbitrators, and the impact on subsequent recall rates and CDR. This interview was undertaken with the Breast Radiologist, who had presented data from a regional 5-year arbitration study (Symposium Mammographicum 2016). The Radiologist also sits on the NBSS programme board, is the secretary to the BIG 18 radiology group and the NHSBSP Clinical Advisory Group (Radiologist 3).

3. The role of future technology (AI) in breast screening. This interview was conducted with the Clinical Director of a Med Tech company (Consultant Radiologist) with research interests in AI as applied to medical imaging and a member of the Royal College of Radiologists Informatics Committee & AI Working Group (Radiologist 4).

The number of interviewees required was difficult to ascertain as the researcher is studying an area that has not been previously explored. Trotter (2012) defines an adequate sample size as sufficient when no new findings are revealed, a concept of data saturation. Guest et al. (2006) suggest that this occurs within the first twelve interviews and that from the initial six interviews, basic meta-themes are evident. The proposed 21 semi-structured interviews were therefore deemed practical with the recognition that this might change during data collection and analysis.

8.2.7 Data Collection Tool

The interview schedule was comprised of questions that would produce valuable and meaningful data. A semi-structured interview guide was developed, which comprised of eight primary open-ended questions tailored to explore the survey

responses further, with follow-up prompts to advance the enquiry. These questions were developed following analysis of the literature review and survey findings and revised after piloting. The interview guide was pilot tested with a research fellow to test question comprehension, interview flow and to provide feedback on the researcher's interview technique. An adaptation to the interview technique following feedback gave the participant time to think and allowed for silence when there was a prolonged pause.

The pilot interview was quite lengthy (timed at 45 minutes), so a decision was made to omit the question on the role of future technology as it was felt this would be better answered from the emergent literature and interview with the Clinical Director (Consultant Radiologist) of a Med Tech company. The final semi-structured interview guide is presented in Appendix 10; questions were structured around (i) reporting and arbitration practices and influences on decision-making, (ii) receptiveness to change, and (iii) opinions and implementation of the PHE arbitration guidance. The interview schedule was flexible so that the order of questions could be adapted depending upon the responses, and additional prompts or questions utilised.

Kvale and Brinkmann (2014) describe nine types of questions that can be utilised when undertaking semi-structured interviews. For this research, a combination of styles was employed in an attempt to elicit adequately detailed information. All nine types of interview questions were used in this study and examples are provided in Table 43.

Table 43. Varying Types of Interview Questions (Adapted from Kvale and Brinkmann 2014)

Type of Interview Question	Example used in this study
Introducing questions: The topic is introduced.	'In the survey, you responded ...?'
Follow up questions: Allows the researcher to expand on an interviewee's initial response	'What is your opinion on that...?'
Probing questions: Direct questioning to explore in more detail.	'So why do you think…'
Specifying questions: e.g. Can you explain?	'I am not sure I understand what you mean by that.'
Direct questions: Receive a yes or no response.	'Do you use Radiographers to undertake third reader arbitration?'
Indirect questions: Utilised to gain an interviewee's true belief	'Is that the way you feel too?'
Structuring questions: Progresses the interview on to the next topic, e.g. '	'OK. Moving on to...
Silence: Pauses can give the interviewee time to think and indicate that you would like them to respond	Periods in the recorded interview of silence
Interpreting questions:	Is that because...?'

8.2.8 Interview Process/Informed Consent/Ethical Considerations

Glogowska et al. (2011) and Musselwhite et al. (2007) emphasise specific strategies to support in-depth telephone interviews. These include the value of advanced communications (e.g. letter or e-mail), initial interview communications (e.g. interview scripts), communication of the purpose of the research and the importance of the participant's contribution. Before initiating the telephone interview, potential participants were sent an e-mail thanking them for their contribution, and a participant information sheet (Appendix 6+7) detailing the purpose of the study, ethical approval, and data safeguarding of the interviews.

Staff were invited to contact the researcher if they were still interested in taking part. Following a confirmatory e-mail, interviewees were sent the consent form to return (Appendix 11), a copy of the PHE guidance and interview topic guide. Additional verbal consent was obtained before commencing the interview to confirm that participants were agreeable to their interview being audio-recorded. Participants were informed that recordings would be anonymised and deleted immediately after transcription, but that extracts from their interview might be utilised in publications, with any identifying names or places removed but job titles remaining (these would not allow identification).

8.2.9 Interview Process

All interviews were conducted on a 1:1 basis by the researcher, who is a female Consultant Radiographer in Breast Imaging. Since the researcher was a novice to qualitative data collection and analysis, a formal training course was completed (23 & 24 May 2017) with the International Institute for Qualitative Methodology (IIQM) (an interdisciplinary institute based in Canada). This was followed by a university interview simulation and feedback session, an initial pilot interview, and supervisory feedback on the first three formal interviews.

Interviews were undertaken between December 2017 and March 2018. Participants were allowed to select a day and time convenient to them. All interviews were conducted during the working week (Monday-Friday), with the majority of subjects telephoned at their workplace and four at home. Four interviews had to be re-scheduled because clinical commitments ran over, or the interviewee had been assigned to a different workplace location, or they had forgotten the agreed date

and time. A unique study code was assigned to each interview transcript to ensure confidentiality. The interviews ranged from 29 to 50 minutes in length.

8.2.10 Reflexivity

Insider Research

It is important to acknowledge the role of the researcher as an 'insider' in this study. There are contradictory beliefs about the insider/outsider role in the research process (Finefter-Rosenbluh 2017). Brannick and Coghlan (2007) assert that there is an inherent bias in any research, but that an insider-researcher has the advantage of inherent knowledge relating to issues of current relevance that an outsider-researcher does not possess. This may enhance the collection and analysis of data, providing the researcher remains reflexive and reflective during the process.

Conversely, insider-researchers may find it challenging to separate personal experiences from participant experiences and therefore, difficult to provide an impartial point of view (Chawla-Duggan 2007). Thus, researchers are encouraged to be reflexive, acknowledging their social position and the impact that their experiences and knowledge might have on the research process (Berger 2015). This is important when the researcher is embedded in the clinical setting (Sim and Wright 2000). It is evident that in this study, the researcher shaped the design; the selection of interviewees; the questions asked, and the issues probed. During the process of sampling for interviews, the researcher could have focused on Consultant Radiographers, peers similar to themselves in some respects. Having a sampling frame with pre-determined characteristics for maximum diversity helped to address this potential bias, and subsequently, only four Consultant Radiographers were

interviewed with a higher proportion of Radiologists (eleven). Two of the interviewees were known to the researcher; however, excluding these professionals would have limited the sampling frame because there were no comparable alternatives. To counteract any potential bias, the researcher remained formal and followed the standard procedural guidelines.

During the interview, individual participants might digress from the question asked, e.g. variation in how to report a digital mammogram. In these cases, the researcher guided the participant back to the question, while being conscious that interviewees' descriptions are based on their life experiences (Todres and Galvin 2012). As Holt (2010: pg 118) states:

"The success of the telephone narrative interview is likely to depend on the telephone skills of the researched as well as the researcher."

Confirmation of any discussion was routinely terminated by using the term 'okay' as 'yes' may have inferred an agreement with the participant's views and influenced them into thinking that this was what the researcher wanted to hear. At times, it was difficult when the researcher was asked what happened in their own practice as it felt like the interviewee was asking for advice. This was addressed by stating that the researcher had experienced a variety of third reader arbitration and group consensus processes while working in various departments. Care was taken not to express personal comments for, or against, these processes. The researcher sometimes felt uncomfortable when individual interviewees (Radiologists and

Radiographers) expressed strong dissatisfaction with a lack of respect in their teams. A Director had also recently left a unit and expressed

> "*They had paid a personal price for trying to maintain an excellent quality service with insufficient resources"*.

In such a situation, where a Radiologist openly conveyed these feelings, the researcher acknowledged that the working environment as described was difficult for the individual so that they would continue to give an open and honest account of their experience, but remained detached from any discussion of such organisational dynamics to avoid influencing.

Tact and neutrality were also essential when interviewing the Directors, who are senior to the researcher, to ascertain why they had not or would not implement Radiographer arbitration/lead of consensus in their unit. This was important so that the Directors did not feel the interview was a questioning of their authority, but rather establishing what factors had influenced their decision. Similarly, when interviewing Advanced Practitioners, the researcher needed to guard against interviewees trying to provide what they perceived to be a correct response. Listening to the audio-recordings provided an effective means of reflection on the researcher's interview technique and the quality of the data produced. The interview style was refined throughout, and this generated more confidence in subsequent interviews.

8.2.11 Qualitative Analysis

Thematic analysis (TA) was chosen as the method for analysing the interview transcripts. This research method is used across diverse epistemologies and defined as

"A method for identifying, analysing, organising, describing, and reporting themes."

(Clarke and Braun 2013)

It was considered appropriate for the explanatory nature of the interview study. TA is a recursive process, allowing the researcher the flexibility to revisit codes and themes to provide a comprehensive account of the data. The 6-phase guide for TA described by Clarke and Braun (2013) was followed in the current study. Themes were developed both inductively from the participant interviews (experiences and opinions) and deductively from the literature. Independent coding was undertaken by a second researcher on a sample of the interviews to verify the reliability.

Phase 1: Familiarisation with the data

Transcription

Interviews were digitally recorded and transcribed verbatim by the researcher for analysis. Bird (2005) states that transcription is a crucial phase for the researcher as interpretation commences at this preliminary stage. Nonverbal communication (e.g. laughter) was also noted within the text. Transcripts were supplemented with notes and researcher perceptions taken during and immediately after the interview (see Table 44).

To facilitate familiarisation with the complete data set, transcripts were re-read, and the audio recordings listened back multiple times. Initial impressions were recorded, for example, where interviewees expressed strong or opposing views. In this study, the main conflicting views were around the centralisation of arbitration services and the appropriateness of complete electronic reporting on NBSS. Familiarisation via listening, reading and note-making supported the researcher in retrieving information from pages of transcript during analysis.

Phase 2: Generating initial codes

The researcher coded each transcript individually using the CAQDAS package NVivo version 11 (QSR International). Each transcript was methodically evaluated. Any interesting or relevant sections were highlighted to describe the content of each passage with a preliminary code, defined by Saldaña (2015) as first cycle coding. Passages of text were often assigned more than one code. Table 44 represents an extract of open coding in which the participant (a Radiologist) talks about how non-blind reading affects the second reader's decision-making and subsequent personal performance. The primary coding labelled this as 'conformity of practice' and notes attached to how the current reporting system does not support fully blind reading.

Table 44. An Extract of Open Coding, with the Researcher's Initial Perceptions.

Coding Labels	**Radiologist 3**	**Notes and Ideas**
Non-blind reading Following-on Clinical Decision-Making Performance Metrics Experienced Reader	If you can't see what the first reader has said and you're acting truly independently I think people would be less likely to conform with what the first reader has said and I know that's certainly true of me when I'm reading I'm much more likely to recall a case that the first reader has recalled than I am to recall a case without an opinion. So, my first reading recall rate is lower than my second reading recall rate. And that applies to a large number of readers even people like me who've got quite a bit of experience.	Non-blind reading: Conformity of readers – negates the purpose of double reading. ? Fully blind reading more accurate reader profiles. Notes – technology, IT systems prevent fully blind reading

Memos were used to record a more detailed note or idea, for example, questions to consider as the analysis progressed, or thoughts on consistencies or inconsistencies within the data (Table 45).

Table 45. Example of a Memo Recorded During the Analysis of Interview Transcripts.

MEMO: 'Technology of NBSS'
Definition
Ideology versus practicality: Blind reading is professed as desirable in the main but challenging to achieve. Practical and safety difficulties (e.g. paper system supporting an electronic reporting system, paper as a failsafe mechanism) challenge the philosophical ideology underpinning a paperless system (complete electronic reporting).
Codes
Non-blind reading; Tension for change; Paper system as a failsafe
The ideology of becoming paperless was contrasted with the practical difficulties and safety aspect of ensuring the right results process. So, although a complete electronic reporting system was philosophically presented as the way forward, the process of actually setting this up in terms of IT infrastructure (NBSS) was seen as far more challenging.

A one in five sample (four transcripts) of interviews were coded by an independent researcher to verify the reliability of the coding and assess concordance or discordance. After these transcripts had been open coded, the labels allocated to the text were discussed in terms of why they were construed noteworthy, what they revealed about interviewees' beliefs and how they might be informative to answering the research question. Generally, the same passages were highlighted as significant. However, there were occasional differences in the language used to

express the interpretation, although the same principle was recognised. Following this stage, the researcher then independently coded the remaining transcripts, noting any new impressions which did not fit existing codes.

Phase 3: Searching for themes

During this stage, 64 nodes identified were reviewed and grouped into comparable categories. Initial codes were refined with duplications removed and discarded if they only contained one comment, e.g. variance in the use of digital tools to report mammograms. The next stage of the analysis grouped the remaining 20 codes together if conceptually related, into seven preliminary themes from which significant patterns could be perceived. NVivo hierarchical charts (Figure 55 and Figure 56), were used during this analysis to visualise how the separate codes were associated, connections between themes, and subsequently to create overarching and sub-themes.

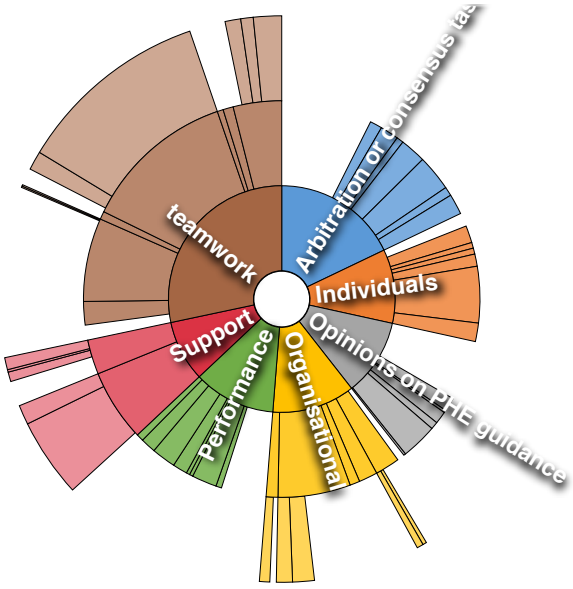


Figure 55. Sunburst Produced in NVivo to Visualize and Compare Data and Themes.

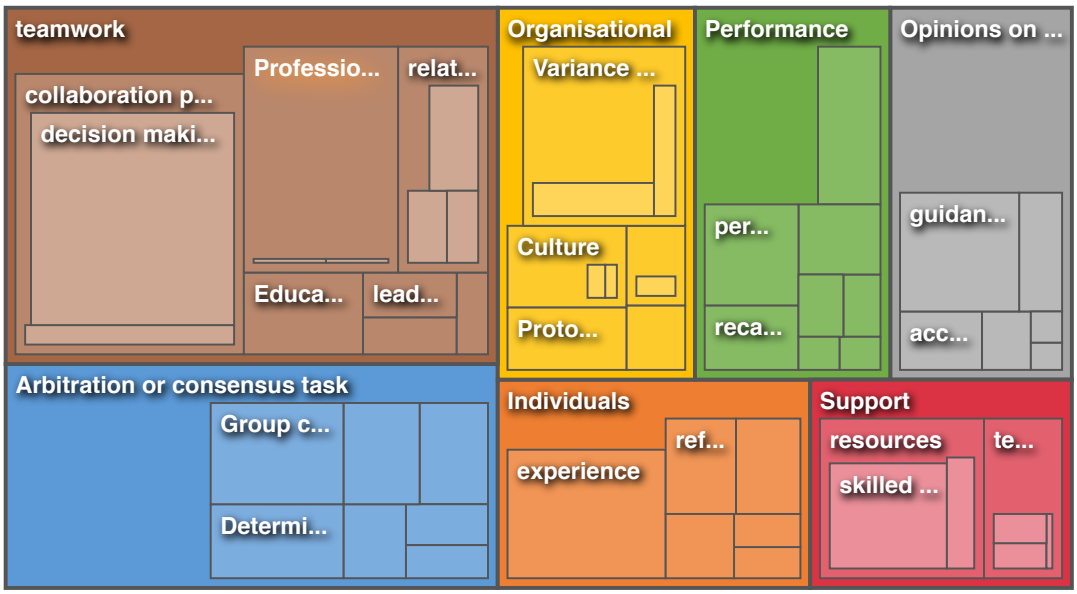


Figure 56. Hierarchical Chart Produced in NVivo Demonstrating how Data were Categorised and the Sub-Themes that Contributed to the Primary Themes.

The interpretation phase involved developing preliminary themes which offered possible justification for what was occurring within the data, and not merely a narrative of individual cases.

Phase 4: Review of preliminary themes

The fourth phase involved the refinement of preliminary themes. The NVivo software enabled search and retrieval by queries and visualisation tools; and cross-tabulation of the mixed methods data. This facilitated practical in-depth analysis of large amounts of data to explore associations between study findings and realise new directions of investigation. Subsequently, some data extracts (quotes) were transferred into an alternative theme. Often, excerpts could have been allocated to several categories which reflects the inter-relatedness of some themes, e.g. non-blind reading, limited by technology but mainly an organisational decision. After this

analysis of the twenty-one interviews was completed, it was judged the data was of sufficient breadth and depth to address the research question (Jolley 2013), and that the thematic map adequately portrayed the data.

Phase 5: Defining and naming themes

Phase five entailed naming the final five main super-ordinate themes and producing an overview which reflected their essence. The narrative needed to depict the participants' views and the principle of the theme concisely and accurately. A table of themes, sub-themes and codes are found in Appendix 12.

Phase 6: Producing the report

The final phase entailed summarising the interviews in a succinct, consistent and rational narrative. Quotes were selected to capture the inherent nature of each theme. It was important that the narrative was not merely descriptive but supported the analysis for each research question.

The next section discusses the findings from the data analysis of the semi-structured telephone interviews.

8.3 Results-Interviews

Following data analysis, the five main super-ordinate themes generated related to organisational factors, technology, clinician factors, teamwork factors and PHE guidance factors, within which there were a number of significant sub-themes (Table 46). A sub-theme was deemed significant if it occurred more than once in the data

analysis, or it was likely to have an important impact on the future provision of the service.

Table 46. Thematic Categories and Associated Sub-Themes from In-Depth Interviews

Super-Ordinate Theme	*Descriptor*	Sub-themes
1 Organisational factors	*Historic cultural elements of reporting and arbitration practices*	1.1 Organisational variance and historical, cultural elements
		1.2 Conformity of practice
		1.3 Silo working and the concept of centralisation
2 Technology	*Information Technology and infrastructure required to support the breast screening system*	2.1 Lack of sophistication of technology to support: • Blind reading and a paperless system • Audit
		2.2 Challenges and prospects of Artificial Intelligence
3 Clinician factors	*Performance measures and associated clinician factors*	3.1 Meaningful measures of performance
		3.2 Difficulties in defining quantitative guidelines for arbitration/selecting individuals
		3.3 Radiographer self-efficacy

Super-Ordinate Theme	*Descriptor*	Sub-themes
		3.4 Outcome expectancy and Radiographer training
		3.5 Decision-making skills
4 Teamwork factors	*Factors associated with teams which inhibit or facilitate group consensus*	4.1 Collaborative working
		4.2 Team dynamics and collegial conflict
		4.3 Accountability
5 PHE Guidance Factors	*Factors associated with the PHE arbitration guidance which inhibit or facilitate the implementation*	5.1 Guideline factors • Evidence strength and quality • Clarity of the guidance
		5.2 Individual professional factors • Lack of agreement • Inertia of practice • Appropriateness • Implementation climate/capacity for change

8.3.1 Theme 1 - Organisational Factors

This theme builds on the survey responses which demonstrated disparate practices relating to reporting and arbitration and incorporates participant's views on the organisational structures that exist within breast screening services. Three sub-themes identified were organisational variance and historical, cultural elements; conformity of practice; silo working and the concept of centralisation.

8.3.1.1 Sub-Theme 1.1: Organisational Variance and Historic Cultural Elements

Although breast screening is a national system, the surveys had highlighted organisational variance in all elements of reporting and arbitration practices. This was also corroborated by the interviewees.

"You think how can it all be different when we are basically doing the same thing, or you think we would, but there's very big differences between places – odd. Having been to (name redacted) for a week and seeing that they do like partnered arbitration and then other places do totally other things. It seems odd to me that we don't all do the same thing but anyway" Radiologist 2

One Radiologist who undertook regional QA visits stated how the number of approaches to arbitration in the country is 'remarkable' and how the lack of standardisation made them feel uneasy.

"It inherently makes me twitch whenever people start doing different things different ways. Although there might be perfectly valid reasons for it that they will argue strongly for and it doesn't strike me that necessarily the patient or the client

is going to understand why we have 100 different ways of dealing with difficult cases" Radiologist 3

A former Director of the NHS Cancer Screening Programmes confirmed that double reading had 'just evolved' resulting in considerable variation. There was limited evidence that double reading was better than single reading (Blanks, Wallis, and Moss 1998), and no further comparison studies had been undertaken.

"We couldn't really say that method A is better than method B with the evidence that we had at the time. And because we didn't have any evidence, and everybody was doing it differently and obviously liked what they were doing, we just left it and thinking well we will just have to come back to that when we've got some more evidence about which is the right way to do it and I don't know whether that time has come" Former Director of the NHS Cancer Screening Programmes.

Participants corroborated survey responses in that organisational variance in reporting and arbitration were related to historical, cultural elements (Appendix 13 Table 47-1A). It is recognised that units need to be allowed some local initiatives, but this has subsequently led to a mushrooming of approaches.

In accordance with some of the free text survey responses (Chapter 6), Table 47/1A (Appendix 13) shows that several participants viewed non-blind reading as favourable, professing that it represents a learning opportunity with educational benefits, enabling the second reader to consider the first reader's thought processes, and offering reassurance. Non-blind reading was also felt to be constructive, as reporting is not a contest between readers. One interviewee considered that blind

reading placed greater reliance on administrative personnel to segregate arbitration cases, rather than the reporters, leading to concerns.

8.3.1.2 Sub-Theme 1.2: Conformity of Practice

This was the most frequent sub-theme identified. The majority of interviewees substantiated survey responses that non-blind reading can influence the second reader's decision-making, and many interviewees depicted this as a negative influence (Appendix 13 Table 47-1B). This situation was exacerbated in two units where the 1st reader annotated on the images the area of concern, which impelled the second reader to recall.

Although one Radiologist described no effect on their performance statistics of whether they first or second read, others (Appendix 13 Table 47-1B) stated that FRQA data demonstrated that their outcomes (recall rates/CDR) were different.

"Using FRQA you can see looking across various different regions and by and large the 2nd reader will find more cancers than the first ….it can't just be chance it must be that some of the cancers they find are you know using the advantage of having yet another opinion available to them not just theirs" Radiologist 3

Interviewees from all professional roles honestly affirmed that there were cancers they would have missed as a second reader (if blind read) had they not got the opportunity to review the images and change their opinion at the time of reporting (Appendix 13 Table 47-1B). One Director considered that introducing blind reading

would cause extra work as the cases missed by the second reader would then be an arbitration case.

This concept of 'following-on' was deemed to be particularly the case for Radiographers and less experienced staff, with the view that some cases would not have been recalled if truly independently read (Appendix 13 Table 47-1B). However, an experienced Director also portrayed it as challenging to disregard someone else's opinion and remain objective with non-blinding reading and arbitration.

"It's really difficult to unlock in your mind the fact that somebody else is concerned about it" Radiologist 3

Significantly, two Advanced Practitioners perceived that non-blind reading incurred a potential for error as readers may be focussed by the first reader's recall and possibly miss further foci of disease or a contralateral abnormality (Appendix 13 Table 47-1B). Interestingly, one Director described how Consultant Radiographers within that unit modified their reading practices dependent upon knowing which colleagues they would be reporting against; anticipating a recall by the second reader (Appendix 13 Table 47-1B). In this setting, they learnt how colleagues read and altered their reading practices rather than developing independent thinking. The Director speculated as to whether this also applied to new radiology Consultants but that their confidence in decision-making is reinforced with feedback from working up cases at assessment; a task not usually performed by Advanced Practitioners. It is important to note that Consultant Radiographers do not run assessment clinics in this unit, but this is not the norm nationally.

Three interviewees (Director, Radiologist and Advanced Practitioner) revealed that this subjectivity was not limited to reading practices, but that third reader arbitrators may also be influenced in their decision-making by knowing the names of the individuals who initially reported the mammogram. Rather than voicing an independent opinion on a potential abnormality and arbitrating effectively, they conveyed their judgement was based on knowing the professional who recalled the case.

"I tend to be selective on who can arbitrate because when they're telling me that because of the reader that's reading, they'd recall or even if they disagreed because they have more experience that isn't arbitration then" Director 3

One Radiologist highlighted how this bias could be detrimental with cancers dismissed as the credibility of high recall readers is undermined (Appendix 13 Table 47-1B). The question remains, therefore, whether NBSS should anonymise the readers with unique codes rather than the individual's initials.

Overall, there were contrasting views on complete blind reading, with some advocates and some in opposition.

"I would be much happier with a totally blinded system I would rather do it without any knowledge of what somebody else has recalled at all. I think it makes you better in the long run "Radiologist 2

"I wouldn't be in favour of that because I actually personally like to have a look and see if something, I might have missed something" Director 6

It is acknowledged that screen reading is difficult, as discussed in Chapter 2. Inherently individuals will miss cancer cases, some of which will present as interval cancers. Interviewees in all professional roles verified that film readers would perform alike, as local culture and systems influence reporting practices, and this is exacerbated when readers have been taught in-house (Appendix 13 Table 47-1B).

"I think what's difficult for us actually as a unit is that most of our film readers are home grown if you like most of our Consultants have come through this unit and have been trained here and all our film readers have been trained here, we're all very similar. So, we all report in a similar way" Director 2

The caveat to blind reading is that it may generate more arbitration cases. Interviewees currently reporting non-blinded surmised there would be an increase in discrepant reads, and this was supported by units that had implemented blind reading (Appendix 13 Table 47-1B). This supports the theory of biased perspective when non- blind reading.

"We had a significant increase in the number of arbitrations when we turned it over to blind reading" Consultant Radiographer 2

The entire purpose of double reading in breast screening is to obtain two independent opinions, and it appears that in some units, the non-blinded reading may negate the process. One Director described this as promoting mediocrity, and that the result is homogenising reading.

"I think that's important it (non-blind reading) does seem to take away like 90% of the point of having two readers really" Director 5

A further consideration raised by one Consultant Radiographer was the belief that blind reading offered greater justification for both readers if they have independently concluded a case normal, that subsequently is confirmed cancer; and may require disclosure of audit.

"The thing that I feel quite strongly about now is duty of candour. That I think it's a much much stronger argument for two people to have reached that decision completely independently to routine recall. I think it's a much stronger argument for the duty of candour" Consultant Radiographer 2

8.3.1.3 Sub-Theme 1.3: Silo Working and Centralisation of Services

The surveys had identified that logistically some units are unable to undertake consensus review due to split site working, current workload and staff shortages. This was explained in greater detail in the semi-structured interviews, for example:

"We do arbitration because of workload. I mean some units do consensus, but you know there is just no way in the working week we could get enough people together to do consensus" Breast Clinician 1

"We don't do consensus because we are split site, so that's the reason that we don't do it" Director 3

With a diminishing workforce and loss of experienced staff, there will be increased pressures on the system. The concept of centralisation of arbitration (but internal to NHSBSP) was explored in the interviews. A significant sub-theme identified by the medics (Radiologists, Directors and Breast Clinicians) was that because services operate in silos, exposure to other reporting practices and external arbitration could be constructive (Appendix 13 Table 47-1C), with the need to work in partnership and support colleagues as desirable.

"I think you know the more you can interact with other adjacent units or they don't have to be adjacent the better, it has to be a good thing I think" Locum Radiologist 1

A significant finding identified in this research is the psychological element of knowing who has reported or recalled a case, and that remote arbitration would provide an independent view. Cases would be judged entirely on the images, resulting in less bias and possibly improved results. Opinions were mixed on external arbitration with interviewees either firmly in favour or opposed. Those interviewees who supported the principle of remote electronic arbitration felt that pooling resources was one of the solutions in making the system viable going forward; especially for the units with staffing issues. Collaboration with other units was considered to allow benchmarking of practices and beneficial to the programme as a whole. Centralisation or co-arrangements with a neighbouring unit was also considered to offer safer clinical governance on arbitration cases and thought favourable to 'fighting' and 'trying' to employ locum Radiologists.

"I think centralised arbitration would be good, I think, especially for outliers it gives you an idea of what everybody else is doing as well or what' s acceptable to people" Radiologist 2

"I would say that offers much better clinical governance than trying to make the decision with one or two people, maybe one of which may have already read the films themselves" Director 6

A future requirement would be a robust electronic infrastructure to support remote arbitration, and the system would need to provide all the information required to make an adequate read (for example, previous surgery) including the availability of previous films. Many units will now have two screening rounds of digital imaging and therefore reviewing of analogue images may only be problematic if a woman had not attended the last two screening rounds. The main issues raised related to the logistics of tracking cases.

"We should be able to do it because it is all image-based, computer-based it should be very amenable to being able to do it remotely and having an independent external third reader would be a very very viable solution I would definitely be for that" Director 5

"I think that's the way we are going to have to go in the future, it's the organisation and funding of that, you know pump-priming and piloting it all of those things" Locum Radiologist 1

Five participants felt that due to the national shortage in breast radiology, pooling of resources should not be constrained to arbitration but should be considered for reporting of batches of films; either on a regional or country-wide basis (Appendix 13 Table 47-1C). It was also suggested that pooling of the film reading from four or five small units would be beneficial in improving outcomes as this increases the diversity of readers, for example:

"I think in this modern age that's the kind of thing I think it will happen not only for arbitration but for first and second reads. So, I think doing regional reporting; country reporting is possible providing that the infrastructure within the trust can support it" Director 3

Two interviewees confirmed that their large unit is currently undertaking reporting for a remote service, and if an assessment is required, this is undertaken at the original screening location. Therefore, remote reading is possible and thus supports the opportunity of centralisation of arbitration and that the IT issues are not insurmountable(NHS England 2019).

"Well actually we do film reading for (name removed – remote unit) and although we had a bit of hassle setting it up in the beginning, it works like clockwork now, so we just read them alongside ours. They do the assessments there but yes its perfectly doable to read them remotely" Director 5

The East Midlands Radiology Consortium (EMRAD) pioneered a digital radiology system comprising of seven NHS trusts (eleven hospitals) with a cloud-based image-

sharing system. This allows the NHS hospitals to easily and promptly share diagnostic images. This revolutionary system has set a

"national benchmark for a new model of clinical collaboration within radiology services in the NHS" (Morley 2019)

and demonstrates that new models of service delivery for the NHSBSP are a possibility. Although one Director had no objection with sending arbitration cases to services within their region, they felt uncomfortable about them going outside. Two Consultant Radiographers felt it was important that arbitrators for the NHSBSP must actively be fulfilling the criteria defined within the arbitration guidance, which may exempt partially retired Radiologists.

"I can say this because it's going to be anonymous because I do know Consultant Radiologists that possibly they're active in the screening service, but they sort of retire part-time or whatever, but they still maintain their private practice. But actually, they're possibly not doing the numbers that they were doing. So, I think you'd have to have very strict controls on that" Consultant Radiographer 1

On the other hand, five staff at all levels voiced opposition to the centralisation of arbitration (Appendix 13 Table 47-1C). They conveyed it may be difficult for staff to accept an opinion from an external arbitrator, especially when there was no opportunity to discuss and rationalise the decision. These interviewees regarded centralising arbitration services as a lost opportunity to learn, a lack of trust in an

unknown third reader, and a sense of inferiority. Segregating arbitration was considered by one Director to be detrimental and a barrier to communication.

"Oh gosh no I would hate that. Well, we are a very well-performing unit and some centres nearby are not so well-performing. When I used to go to the interval cancer meetings, we would listen to the presentations of the missed assessments, and I would be horrified sometimes at what they were doing. So, to have somebody else and it's like a control, I would hate it. I think we're all control freaks anyway, but I wouldn't trust somebody else to do it" Advanced Practitioner 3

Communication between the external arbitrator and assessing team was considered crucial to ensure the correct area would be worked-up at assessment, and there was an apprehension that another area of concern might be detected. However, this situation can arise with in-house third-person arbitration and review of cases at assessment due to the subjective nature of breast reporting. With digital technology, it is possible to annotate the area of concern on the mammograms, and this may be revealed after the images have been preliminarily viewed to avoid biasing the reporter. As discussed previously if current practice is homogenising reading in some units, external views could be considered educational. If an individual required to work-up a case at assessment disagreed with an externally arbitrated view to recall and the abnormality was subsequently proven to be cancer, they will acquire knowledge that is unlikely to be derived in-house. One individual expressed the view that staff may lose the skill set and expertise in making decisions on difficult cases.

"It's absolutely that I think we are going to have to look at different ways of delivering the service and that is most certainly an option and I think it's a realistic option for a good many units, but you don't want to take the expertise out of the bigger picture either do you?" Consultant Radiographer 2

A further concern raised with external arbitration was the potential of introducing an inherent delay into a programme that is tightly governed by time constraints (Appendix 13 Table 47-1C). This would therefore need to be a reasonably rapid service to avoid external arbitration being counterproductive. One Radiologist voiced that it is inappropriate to oppose external arbitration merely because individuals do not feel comfortable with it. They advocated that the outcome for the women is the priority and if a system gives better results, is practical and affordable, it should be considered. External arbitration may potentially reduce the variation in recall rates that exists from differences across units, driving improved quality through sharing and standardisation. It may also offer the opportunity to maximise workforce expertise.

"It sort of standardises practice a little bit doesn't it if you send yours to an external third reader then you have an idea about what other people thinks acceptable to call back or not to call back" Director 5

It is recognised that this might require a change in culture and mind-set, that would necessitate sensitive handling to avoid people feeling threatened. The former Director of the NHS Cancer Screening Programmes considered centralisation of

arbitration was achievable and akin to the French system for second reading; albeit a revolution that would require considerable investment.

"I think you could do that and it's very similar to what the French do, second reading is done centrally at a later date. But for us, I think we would have to have everything electronic and all the images held centrally, that's how we would have to do it, which is you know possible but it would be a revolution and would take a lot of investment, but then it could be done. I think it is a possible solution, but you would have to have the electronic infrastructure to support it" Former Director of the NHS Cancer Screening Programmes.

8.3.2 Theme 2 - Technology

Results from the surveys showed only a small number of units (n=4) reporting non-blinded (14 units partial blind reading). This theme builds on the survey responses and incorporates participant's views on the information technology and infrastructure required to support reporting and arbitration practices in the future. Two sub-themes identified were a lack of sophistication of the current technology to support blind reading (paperless system) and audit; and the challenges and prospects of using Artificial Intelligence in breast screening.

8.3.2.1 Sub-Theme 2.1: Lack of Sophistication of the Current Technology to Support:

- **Blind Reading (paperless system)**

A common theme identified via the surveys was that the current NBSS reporting system is heavily dependent on paperwork. Four interviewees (Radiologist,

Directors and Consultant Radiographer) described the paperwork as onerous and preventing complete blind reading (Appendix 13 Table 47-2A).

"Why are we filling out paper assessments though, it makes no sense to me at all we should be totally computerised. I would be much happier with a totally blinded system which has no paperwork to it either" Radiologist 2

A change in the reporting software (NBSS) would be required to facilitate a paperless system, and hence opinions were explored regarding the pros and cons of developing an electronic proforma that would automatically be generated if a recall was selected. While the majority of staff from all professional roles supported a move to a paperless system which would support a more objective approach for the second reader (Appendix 13 Table 47-2A), there were polarised opinions.

"That's part of our work here because we are trying to go paperless within our own department and NBSS stops us" Advanced Practitioner 2

The paperwork was deemed to create inefficiency. One Director that undertakes QA visits stated that some units require the reporters to mark the paperwork regardless of whether they consider the case normal or abnormal. Concentrating on clerical tasks amidst reporting was considered a distraction and reporting normal cases should be seamless. There was clearly a subconscious distinction when recording (currently write) information for the recall cases. The system needs to support the reporter in making the clinical decisions rather than being burdened with paperwork tasks. A fully electronic system would negate the need to wait for paperwork to

arrive from the screening mobiles and would support the option of remote reporting and arbitration as discussed previously.

Three interviewees (Director, Radiologist and Consultant Radiographer) were opposed to a fully electronic system as blind reading removed the ability to go back and review the 1st reader decisions. Also, they believed an electronic proforma would increase their reading time. They preferred paper as they 'liked to draw' and considered it easier to convey and understand a colleagues reasoning in a written format.

"If there was no paperwork at all that makes it completely blind and I wouldn't be in favour of that because I actually personally like to have a look and see if something, I might have missed something" Director 6

"I think that will add time to reading is my concern" Radiologist 1

They were also not in favour of an electronic proforma as the current diagrams within the assessment section of the NBSS system are 'not ideal' and a 'bit clunky'. The breast diagrams are split into squares, and therefore it does not allow the reporter to document a lesion at the 12, 3, 6 and 9 o'clock position. A concern was raised over whether there could be an interpretation error of the correct area to assess if using an electronic proforma compared to paper. In the researcher's practice after the discrepant images have been reviewed at consensus, and there is a decision to recall, the area of concern is digitally marked (circled) and saved as a separate image to the Picture Archiving and Communication System (PACS).

Therefore, it is clear for the responsible assessor the area which has been reviewed and discussed by the consensus team.

There is a requirement to be able to access essential information regarding previous surgeries, scar sites, patient's symptoms etc. However, one Advanced Practitioner commented that currently, it is uncertain whether a reporter has recognised and reviewed the written comments on the paper proforma (Appendix 13 Table 47-2A). The system has an electronic alert that identifies relevant symptoms, for example, complaining of a lump, that the Radiographers enter at the time of the screening mammogram. Any alerts must not be distracting to the reporting workflow but easily accessible.

Overall, the NBSS system was deemed to lack sophistication, but it was recognised that there was scope for developing an electronic proforma if funding for NBSS rewrites can be secured. The development would, therefore, require breast screening reporter input and must be user-friendly.

"We've all been saying it needs to be a pie, a clock face you know type thing. I think NBSS online I'm afraid to say is a very, we all know very outdated computer system that's been updated, but really if we had the money, we'd scrap it and start afresh"
Radiologist 1

However, the counterargument identified with entering a recall electronically was the lack of paper as a failsafe mechanism. Staff from all professional roles thought that an error could be made by selecting the wrong outcome on NBSS. i.e. abnormal

when normal and vice versa (Appendix 13 Table 47-2A). Currently, reporters have to actively select the term 'abnormal' from a drop-down list. If an electronic proforma were generated on selecting 'abnormal', this would prevent the potential error of reporters writing the wrong name on a paper record. However, a more significant concern was the ability to select the normal/normal option in error on NBSS when a recall is intended. The current facility to cross-check a paper record with the computer entry allows identification of a discrepancy and potential error.

"If you accidentally put in the wrong patient or something you've got two methods of flagging up the abnormal ones so you can't possibly miss anybody. So, if you accidentally write down the wrong person's name on the sheet they are still ticked as abnormal on NBSS so they will still appear in the consensus list and vice versa if you accidentally you know when you're distracted by something, and you write it down on your sheet, and then you just go and click normal/normal by mistake you've got another sort of safety net really" Director 5

With double reading, it is less likely that both readers would make an entry error on the same patient, the only caveat being if only one reader perceived the abnormality and incorrectly entered a normal result.

As raised in the survey responses, participation in specific clinical trials mandates blind reading, and therefore some units are having to make changes to their current working practices and failsafe for the right results. However, this change in practice is met with hesitancy.

"We are moving to blind reading because of the PROSPECTS study. I'm a bit apprehensive about going blinded because I know that that will increase arbitration, but I'm also concerned that the learning element goes and that's what the concern that I have with that" Director3

- **Audit**

The survey responses (Chapter 6) also highlighted that currently, there is no uniform way of recording onto NBSS the individuals present and accountable for the outcome of cases at consensus review. Although this may be evident on a paper trail, capturing the personnel present directly on NBSS would potentially facilitate more accessible audit and possibly improved consensus statistics.

"We enter it under the name Arbi and yeah so that's the only identifier on NBSS, we do have a paper record of individuals who are that session and that's kept, that's kept separately on an A4 sheet, our manager files away I guess, and it's never looked at again I would imagine, but it's there" Locum Radiologist 1

One Director assigned two-digit codes to all readers generating a four-digit code for consensus (first reader's digits and then the second reader's digits). This was a large unit, and hence this generated a vast mix of consensus ID's. The Director described how analysis of this data was 'painful in the extreme'. If NBSS were updated to incorporate an option to select consensus or third reader specifically and enter the individual(s) present (initials or anonymised code), it would provide a more reliable method for data analysis. Amalgamating several years' worth of data would generate more substantial numbers of arbitration cases and facilitate an output

from the pairs/group combinations and allow the analysis of specificity, recall rates, CDR which could inform optimal structuring (staffing levels permitting).

"If NBSS just allowed you to say these people are in the room during this you know during this session then it makes more sense" Director 1

The PHE arbitration guidance states regular audit (personal and team results) and reflective learning as one of the recommended requirements. The importance of audit and feedback in terms of being able to review images and the individual's decision-making is considered imperative if readers are to improve. The crucial factor identified was learning from the review and people changing their practice accordingly. It is acknowledged that it may be difficult discussing how individuals are performing between peers as it is not easy for people to expose their weaknesses to others. The survey responses highlighted that in some units, there are no processes in place for feedback from consensus meetings and interviewees reported varying levels of opportunities to undertake this.

"I think auditing consensus would be, it's something we don't do, and I think it's something we should do because it's really important that we know if patients who've been recalled by one or more readers and then been routine recalled at consensus" Advanced Practitioner 1

With increasing staffing shortages and rising clinical demand, there was a concern that audit, and reflective practice opportunities may diminish. Although units will have systems in place to review interval cancers (including arbitrated interval

cancers) and false-negative assessment's, auditing practice is variable (Appendix 13 Table 47-2A). The ability to collate and review a larger number of cases is more likely to identify if there are particular patterns or individual trends and provide the opportunity for people to change their approach to assessment.

"So, what we need is ideally at the end of the year NBSS would do exactly what I did very manually. Here's the list you press the button it comes up on PACS, these are the ones that pertain to you, but this is all of them. All that data should come up at the press of a finger. You know if we arbitrated it last time was that a little cancer and it's now a bigger cancer. You know NBSS makes it so difficult to do all of that" Director 1

Overall, team members have limited time and supporting tools or the infrastructure to evaluate and reflect on their performance efficiently.

8.3.2.2 Sub-Theme 2.2: Challenges and Prospects of Artificial Intelligence

The combination of 'big data' and Artificial Intelligence (AI), represents a revolution in medical imaging. Radiology must strategically plan for a future in which AI is part of health care delivery. Although this was not explicitly included in the staff interviews, it was raised by two participants with differing opinions, one arguing that AI is not the solution for substituting a second reader, with the other thinking that AI would be a valuable decision-making tool on discrepant cases.

"Maybe you could have CAD arbitration how about that. You could just leave it to a computer to make the final decision" Advanced Practitioner 1

"I know there's experimenting with CAD and single reader and so on, but I am not sure it's the answer. I am sure we can all find cases that would challenge or defeat an AI system you know" Locum Radiologist 1

With the shortage of breast Consultant Radiologists predicted to increase over the next five years, human resources to double read may prove problematic. An interview with the Clinical Director (Consultant Radiologist) of a Med-Tech company was undertaken to understand the challenges and prospects for using AI in the breast screening setting. The interviewee anticipated that deep learning can potentially massively improve the decision-making process as compared to traditional CAD software which is not very specific. The Med Tech company's algorithm has been implemented in hospitals this year, but in a research capacity (18-month Wave 2 Testbed project), rather than clinical practice. The company undertook an independent multi-centre clinical study to evaluate their software's performance before submission for CE marking. The retrospective study results are awaiting publication. Hence, exact figures were not disclosed, but on a case-wise basis, the interviewee stated that the software has a sensitivity and specificity higher than any existing CAD and stronger than a single expert breast Radiologist. The inference is that this particular AI will be able to differentiate benign from malignant findings better than a single human can, but as of yet the system has not been tested against standard UK practice of double reading and consensus. The interviewee considered it realistic that in the UK breast screening programme, we may in the future have single reading supported with AI. Returning to a single read would be worrisome to some staff because of the risk of litigation, and the number

of cancers that are only detected by one of the readers. Double reading confers a safety net.

"Systemically at the end of every year once the KC62 was closed I would pull all my single reader cancers which is about 25% of the cancers" Director 1

"With reading as it stands at the minute, and this might change if they did go back to single reading, is that you are always hidden behind somebody else. It's never a lone decision ever" Consultant Radiographer 2

The concept of using AI as a learning tool, a platform to help individuals improve their reading practice was also explored. It was confirmed that AI could support individuals in either a prospective or retrospective review of cases.

"Absolutely one of our intended uses is quality control and training of Radiologists. So, either you just show me a case you make a decision and then we show you what we thought it was, so that's one way of doing it…… we can look at your last 1000 mammograms, and we can show you how much we agree with you or disagree, or the ones you may potentially have got wrong" Radiologist 4

At the time of the interview obtaining regulatory clearance was described as the next main hurdle (subsequently European CE approved independent second reader, FDA underway) following which collaborative working with clinical partners would be required to test the software. Reporters in the NHSBSP are required to undertake the PERFORMS test (Gale 2010) set of difficult cases. The ability of an AI

system to read these cases would provide interesting results and may instil confidence and trust if results were on par with reporters nationally.

"So, you know it's a long process, all of that stuff, data gathering labelling the data, regulatory clearance, running trials it's three or four years at least" Radiologist 4

8.3.3 Results Theme 3 - Clinician Factors

This theme builds on the survey responses which demonstrated variable professional roles undertaking arbitration and incorporates participant's views on clinician characteristics pertinent to reporting and arbitration practices. Participants were asked what they classified as an 'experienced' reader, how they selected arbitrators, and how to define quantitative guidelines for new arbitrators. The five sub-themes are I) meaningful measures of performance, II) difficulties in defining quantitative guidelines for arbitration/selecting individuals, III) Radiographer self-efficacy, IV) outcome expectancy and Radiographer training and V) decision-making skills.

8.3.3.1 Sub-Theme 3.1: A Meaningful Measure of Performance

Although the PHE arbitration guidance recognises the skills to undertake third reader arbitration, or to coordinate/lead a consensus review, are not necessarily related to the professional role, the quantified metrics relate to film reading numbers.

'reading >= 5000 films per year including 1500 first reads, 4000 screening mammograms'

and experience

'be an experienced film reader > 2 years in breast screening'.

The surveys had highlighted that the classification of 'experienced' was highly variable. This was corroborated by the interviewees. The predominant view was that the phrase 'experienced' is meaningless, particularly regarding years and number of films read. The breadth of exposure was considered an essential factor, which can increase an individual's knowledge in a shorter period. The inference is that there is a requirement to introduce some philosophies where status depends on proven competence. Competency would be defined by an individual's sensitivity/specificity, recall rates and subsequent PPV. Competence rather than professional role and years of experience would better define individuals suitable to undertake third reader arbitration or lead consensus reviews (Appendix 13 Table 47-3A).

"I think that you should get rid of the phrase experienced and inexperienced and say proven, proven level talk about proven levels of sensitivity and specificity to categorise people on what they have actually shown they can do. You know you can be bad 20 years down the line, I might be bad twenty years down the line, I know I'm not particularly good 20 years down the line" Director 1

A Radiologist who was consulted on the draft guidance clarified that the two years of experience stipulated, primarily was a conservative way of stating it should not be 'rookies' undertaking the task. However, it was acknowledged that there might be

exceptional cases. The rationale for two years relates to the view that data for less than two years is considered insufficient to review an individual's performance. One Director voiced a strong exception to the NHSBSP standard for the volume of films required to be read, saying this was excessive, not an indicator of performance, and believing the 'American studies' suggesting 2000 mammograms a year is sufficient.

"Do you want one of my biggest bugbears of all one of my biggest bugbears of all is -volume of films read because it is no measure of a reader's competency and actually because of the problems in the unit I was in I often read 15 to 20 thousand in a year, but I wouldn't read them very well you know I would read them fast and my recall rate was low, but my cancer detection rate was nothing to be proud of. It was fine I was within parameters, but I think volume read is no measure of competency at all and I actually feel worked up about it as you might guess" Director 1

8.3.3.2 Sub-Theme 3.2: Difficulties in Defining Quantitative Guidelines for Arbitration/Selecting Individuals

As discussed in Chapter 2, all readers have weaknesses in screen reading, and therefore this provides justification for distributing single third reader arbitration over as many individuals as possible to minimise personal blind spots. The counterargument is that numbers are then so small that the figures become less significant and less informative. Although third reader statistics can be obtained by running FRQA reports, two participants stated that little consideration is given to them.

"But I never know what to make of the 3rd reader outcome" Radiologist

"Yes, we've got that PPV value anyway in the FRQA don't we for first, second and third reader and the problem is people don't pay much notice to it" Director 3

Third reader statistics are disregarded in BSIS, for a valid reason. The problem in defining quantitative guidance for arbitration is that these are a subset of cases, not the general screening population. Therefore, it is not possible to compare sets of arbitration cases between units as the cases sent for review will depend entirely upon the characteristics of the readers within individual units, and local reporting protocols, for example, recall of well-defined solitary masses.

"If you're working in a unit where you know somebody recalls every single well defined rounded mass be it single or whatever even if they are multiple, if you're then third reading theirs, you're gonna say you know, no to a lot of them, and that's your job as the 3rd reader" Radiologist 1

If there are extremely conforming readers and automatic recall when both readers agree, there will be a minimum number of cases requiring arbitration. Conversely, if there are nonconforming readers, this will generate a more significant amount requiring arbitration. The sensitivity and specificity of the reader combination will also determine if there will be a higher proportion of cancers in the arbitration group. If reader profiles are similar (high or low), there will be a low threshold of cancers at arbitration. Hence, it is challenging to determine how success in

arbitration can be measured. One Radiologist suggested that a way to define it would be to:

"Look in a unit at paired readers, so reader A and reader B and then take out of the pile that they arbitrate, if they are regular arbitrators, all cases where either A and B have had anything to say. If you do that, then you're only dealing with the same cases, in theory, so the mix should be the same, and their Cancer detection then should be the same. If you find one's got a much higher or lower cancer detection than the other, that would tell you something that would be useful" Radiologist 3

An alternative is to review the positive predictive value from arbitration, as there will be a higher proportion of cancers in an arbitration pile compared to a standard reporting batch. Although not directly comparable, the PPV should approximately be the same across all arbitrators and could provide a reference to assess if there is a significant disparity between individuals. However, with individuals only arbitrating small numbers of cases, the statistics can be misleading. One interviewee stated that as linkage of the cancer data registry with the screening history has improved, over time it will be possible to segregate the cancers which have been arbitrated back to routine recall, providing another way to look at performance in arbitration.

"I suppose the objective way of looking at it is to see what their interval cancer rates are isn't it, and in those units that have a higher interval cancer rate you can try and match that to arbitrated cases then you know that might be a way of determining whether it's practice which is deemed acceptable or practice which is outside of what you would expect" Locum Radiologist 1

The interviews highlighted that how sensitive and specific the third reader should be has never been specified, but participants acknowledged that there would be considerable variation between individuals undertaking the task (Appendix 13 Table 47-3B). Subsequently, this impacts on the capacity required for assessment clinics and ultimately, the number of cancers detected. This was supported by three Radiologists describing the data they had reviewed within their region.

"I know from our practice that those of us who arbitrate that the recall rates differ widely for arbitration so that we've got some people who recall about 1/3 of the cases they arbitrate and other people who its nearer to 2/3 and you know somebody else is about 50%. I can't remember them precisely, but there is variation there" Director 2

In a unit with a high prevalent recall rate, the Director described how they actively circulated individual and unit performance data to allow comparison with peers and the review of linear data to evaluate trends. Even though this feedback stabilised recall rates, it did not achieve the desired reduction in recalls. Two interviewees described that their unit had selected arbitrators based on known performance data, which had a dramatic effect in one unit's recall rate (40%). This was stated to initiate a change in all the reader's practice and consequently had a long-term benefit.

"So, all of the prevalent get arbitrated by either a group of people or a single arbitrator but the two single arbitrators that are allowed to do it have been selected based on their first and second reading characteristics" Radiologist 3

Although the performance of readers with high sensitivity and high specificity at first reading may not translate directly to third reader arbitration, BSIS data identifying individuals in that quadrant may potentially inform who the best readers would be to undertake the task. It was considered that, as the BSIS system will provide more information and a visual aid (a graphical representation of sensitivity/specificity) of which quadrant individuals are in, it may encourage people to think about it. This data may also be useful in determining individuals that might not be considered, for example, individuals who may have a low recall rate but also a relatively low cancer detection rate.

"If they are somebody who perhaps has a relatively low cancer detection rate, they maybe have a tendency to normalise things" Director 2

The interviews highlighted that third reader arbitration is a significant element of the service that has had little consideration. One Radiologist stated that it would not necessarily be in the consciousness of most Directors to think about how the arbitration process might be quality assured because the statistics make this a complicated process.

"I think it's hidden, ……but it's in danger of it being a really key part of the service that we don't pay any attention to. There has never been that much work done even when I was inspecting I never did that much work asking the unit Director. In fact, to be honest, it wasn't even one of my questions, it should have been. Asking the unit Director how do you choose who arbitrates are you happy with that? How do you monitor the outcome of that? Radiologist 3

One interviewee suggested a trial of arbitration cases as a means of assessing performance for potential arbitrators. There would need to be the correct incidence of cancer in those groups as performance cannot be judged if it is too artificial. One Radiologist also suggested the more radical proposal that instead of film reading for the breast screening programme, certain individuals become arbitrators. It was considered that true independent arbitration could be a powerful tool for normalising arbitration across the country and would make it possible for sufficient arbitrations to be carried out by one person so that performance could be continuously monitored.

"If your job is actually, I arbitrate for the NHSBSP; you don't read films anymore; you are now an arbitrator, you could then do 5000 arbitrations. That would be really hard work, but the upside of that would be maybe better arbitration, better learning for units, better monitoring of how arbitration works, and a system that just plain works better "Radiologist 3

8.3.3.3 Sub-Theme 3.3: Radiographer Self-Efficacy

The surveys had identified a 23% agreement with the statement 'Radiographers in the unit do not want to undertake this role (third reader/lead consensus)'. This was explored in greater detail in the semi-structured interviews. Interviewees of all professional roles reported Radiographers not feeling confident or wanting the responsibility of undertaking an individual third read (Appendix 13 Table 47-3C). Given the risk of dismissing cancer, and the possibility that such an oversight could

lead to litigation, may explain some Radiographers' reluctance to undertake third reader arbitration or to practice defensively, recalling a higher proportion of cases.

"We allow them to do it. They choose not to do it. …. I think they feel very vulnerable". Breast Clinician 1

In contrast, three Advanced Practitioners showed enthusiasm for the extended role and were convinced in their ability to perform on par with colleagues and make judgements to override a recall and return a woman to routine screening (Appendix 13 Table 47-3C).

"So, I think if we're able to film read and ……. and if your standards are equal to everyone else in that field, you know within your cohort. Then if you're all at the same level, reading at the same level. Why shouldn't you be able to do it?" Advanced Practitioner 2

The concept of a transition period for people new to undertaking third reader arbitration was raised by a Consultant Radiographer and Radiologist, not only for Radiographers but for Radiologists new to breast screening. The ability to actively review outcomes against their judgement had helped a Consultant Radiographer to develop self-confidence in their decision-making on discrepant cases (Appendix 13 Table 47-3C). Although a Director in another unit offered this supportive feedback, the Radiographers opted out.

"When I was Director that was something I did suggest, you know, perhaps they wanted to try it, and put what they thought but leave it back on the arbitration pile

then we would then do the official arbitration, but no they, they have chosen not to do it" Breast Clinician 1

8.3.3.4 Sub-Theme 3.4: Outcome Expectancy and Radiographer Training

Outcome expectancy is defined as an expectation that adhering to guideline recommendations will result in better patient outcomes. The negative comments regarding Radiographer third reader arbitration related to reduced outcome expectancy, in terms of higher recall rates. Radiologists and Directors justified not delegating the task because Radiographers were considered too inexperienced or had unstable or high recall rates (not necessarily associated with higher detection rates) (Appendix 13 Table 47-3D).

"The Consultant Radiographer.....they are now in their third year of film reading, and the Advanced Practitioner one has had more than three years of experience, and the other one is in her second year, so they are relatively inexperienced, but also their recall rates are higher. So, none of them are arbitrating because their recall rates haven't been steady and low over a consistent period "Director 3

Three Directors and a Radiologist highlighted a distinction between Radiographer and medical training, with Radiographer training reliant upon protocol-based practice, risk aversion and subsequently uncertainty in decision-making (Appendix 13 Table 47-3D). This may reflect the disparity in the time required to become an arbitrator expressed by some Radiologists, with two years considered insufficient for an Advanced Practitioner, but not for a Consultant.

"Doctors are trained throughout their training to make decisions and to take responsibility for decisions, and that is the big thing that Radiographers and other paramedical staff generally are not, you know their training is different your training is much more about protocols and following protocols and things. So, there's actually a whole different kind of change in ethos and thinking. So, I don't think two years is enough for you to feel like you'd have enough experience of screen reading to be a third reader. I think I'd want five" Radiologist 1

Radiographers at Advanced and Consultant level acknowledged that they might overcall, and a significant factor appears to be the personality of the individual regardless of their level of advanced practice (Appendix 13 Table 47-3D). The main characteristics related to lack of confidence and assertiveness; which was depicted as a lack of willingness to speak up in consensus or to have 'courage in their convictions'

"Yes, we might end up over calling. That's the danger, but some Consultants might overcall as well" Advanced Practitioner 2

"Consultant mammographers ……. and actually, their role in arbitration ermm depended on their personalities, not the fact that they were Consultant Mammographers. So, you know some of them were good at this, and some of them were not so good at making the decisions" Director 1

In one unit, utilising Radiographer arbitration, a lack of decisiveness predictably resulted in recalls, but this was supported rather than incurring breaches. From the

researcher's perspective, this is questionable as the ramifications of a recall for the individual can be significant, and it appears that achieving the NHSBSP standard was considered the priority. Arbitration is only worthwhile if whoever is undertaking the task is prepared to make negative decisions.

"What was better was it almost inevitably going to be recalled because they would always air on the side of caution an additional recall or to have it breach. You know as a unit we had to weigh up the pros and cons and we would rather that we just kept things moving from the screening point of view than ermm have this poor packet just sitting waiting for you know somebody with the alleged appropriate expertise to be on-site" Director 2

8.3.3.5 Sub-Theme 3.5: Decision-Making Skills

The PHE guidance states that the 'arbitration process requires different competencies to those of film reading', especially 'decision-making skills with good specificity'. This was corroborated by several interviewees who supported the view that it is a different decision-making process to 'sifting through a load of normal mammograms' (Appendix 13 Table 47-3E). It is not only the ability to detect a potential abnormality but, crucially, assessing the likelihood of it being a malignancy. Regardless of the process (third reader arbitration or group review), comments made by three Consultant-level staff related to the film reader's (Advanced Practitioners) lack of appreciation of the assessment process (Appendix 13 Table 47-3E). The implications of decisions they make were expressed, especially to the

women recalled; and in some cases, it was unfeasible to obtain a definitive diagnosis and might subsequently result in short-term recall.

"I know this is where they said that Radiographers can arbitrate provided that they are involved in the assessment setting, and I think that's important because I think unless you know the consequences of what you are recalling I don't think you understand the importance of making the decisions and people would say oh we couldn't ignore that it's a definite abnormality, but when you think about the impact it's had on the lady for a Fibroadenoma or a cyst you're more comfortable saying that doesn't need to come back" Director 3

Therefore, the ability to work-up cases at assessments was viewed as a distinct advantage. This process was deemed to provide gradual assimilation of information and continuous learning. A Consultant Radiographer confirmed that, before undertaking this role, she had a lower threshold for reporting cases as normal/benign. This suggests that there may be a transition of knowledge when working at the responsible assessor level.

"Yes, before I started to do the assessments myself ermm I think I was, I would probably argue more for someone coming back than not coming back" Consultant Radiographer 3

The question remains therefore whether film reading alone, as an area of advanced practice, may not be optimal. Advanced Practitioners who also undertake biopsies

may have a greater appreciation of what they recall, with potentially improved clinical judgements.

One Consultant Radiographer broached the opinion that it was not only Radiographers who needed to be actively undertaking decision-making in the assessment and MDT, but it was perceived that there had been a behaviour change in locum and semi-retired Radiologists. Although they confirmed it was a subjective opinion, they had observed these individuals were recalling more cases and demonstrating more cautious behaviour, which they deemed was due to concern over missing a cancer and being criticised by colleagues.

> *"It's about whoever is doing it. I am not that keen on locums arbitrating; I don't want people who have drifted out of the service arbitrating. I've got to admit this is anecdotal, but I would say their specificity goes down, they start calling more. I would say with the two people that have retired (in inverted commas), I have noticed a change in the way they arbitrate…. they call more back" Consultant Radiographer 2*

Currently, due to staffing shortages, services are utilising staff to undertake third reads/lead consensus who may no longer attend MDT's or actively undertake assessments. Radiographer personalities appear to be a significant factor differentiating individuals who can translate the knowledge and skills acquired into effective clinical judgment and decision-making. A lack of confidence and decisiveness have been identified as issues with Radiographer arbitration. Therefore, this study's findings indicate that to develop Radiographers as autonomous decision-

makers, the training programmes need to ensure they are developing the skills (intellectual and cognitive) required to manage difficult cases and to make definitive judgements.

The concept of aptitude tests was raised by two Radiologists when selecting Radiographers as film readers. There was a suggestion that Loughborough University who provide the PERFORMS test could produce a test to assess a novice's ability in decision-making. There is currently an IMPROVE (Gale 2010) scheme administered through some of the UK breast screening training centres, but this is undertaken after individuals have started their training and then towards the end.

> *"Is there a way somehow of determining that beforehand and maybe if Loughborough could devise a test to assess somebody's screen reading ability from a greenhorn who's not done anything or very little that would maybe be a useful tool if someone wants to go down the path of screen reading as a Radiographer"*
> *Locum Radiologist 1*

Although PERFORMS is an educational self-assessment and training scheme, a Radiologist who has been undertaking breast screening for many years, reported being a soft outlier in the recent test and described a subsequent lack of help, and limited resources for readers to improve. It was considered that the current PERFORMS is used as a measure of performance, with follow up actions for individuals who are considered as under-performers. It was suggested that, alongside the current scheme, a separate set with more education included could optimise reporting skills (Appendix 13 Table 47-3E).

"Maybe the other PERFORMS we do should be an education thing and maybe they could actually concentrate you know actually do one that is heavily weighted to calc or distortion, you're not judged on it, it's for your feedback and then maybe if you got your feedback, and it said you missed particularly the calc then you could access online help for you know reviewing some cases" Radiologist 1

8.3.4 Results Theme 4 - Teamwork Factors

The surveys had demonstrated that staff attitudes towards consensus meetings and third reader arbitration might differ in the organisation. The interview questions encouraged further exploration in order to understand the team dynamics and subsequent effect on decision-making within a consensus group. Three sub-themes are identified from the interviews: collaborative working, team dynamics/collegial conflict and accountability.

8.3.4.1 Sub-Theme 4.1: Collaborative Working

In accordance with several survey responses, interviewees from all professional roles voiced a positive attitude to consensus group review, describing the process as valuable and conducive to learning as shown in Table 47-4A (Appendix 13). In specific units, consensus was portrayed as a collaborative process providing an educational opportunity; allowing discussion of difficult cases with peers, and a means of disseminating experience to team members and forming departmental policies.

"Most of our arbitration is done as a group to make it an educational and so I think it's you know it's an interesting exercise doing arbitration and ermm but it's

time-consuming, (laughter) and to do it as an educational exercise is quite time-consuming but I think that's where you learn the most from it whatever your level of expertise because it's good to hear you know and to look at what other people call back you know, we all know that we've got strengths and weaknesses as film readers" Director 2

Participants from all professional groups highlighted the importance of having an open, supportive, and respectful team in which to voice their opinion (Appendix 13 Table 47-4A). It was considered that consensus meetings should be facilitated rather than led, which would promote equity across a team, an appreciation of everyone's view and subsequently improve teamwork.

"We've got two film readers and only one Radiologist, and we go we don't need that back, he'll say that's fine. You know, and he accepts us for the experience that we have, which is really quite nice. Whereas I know in other units you wouldn't be accepted quite the same, you know" Advanced Practitioner 2

"*We certainly do want to take aboard the opinion of all the film readers we don't just make the decision ourselves if there's anything where it's a little bit uncertain we will ask; I will ask if anybody has a view contrary to what we are entering" Director 6*

It was acknowledged that consensus might also allow poor judgements to be circulated amongst a team (Appendix 13 Table 47-4A). However, it was considered

that attention would be drawn more promptly to inferior judgements broadcast via a consensus approach and therefore be less likely to result in maverick practice, representing a potential point of failure, than might occur with a single third reader arbitrator.

"Then that bad thought might stand out more quickly ermm. So, ermm certainly when I visit places it's really difficult to say to places that don't really have much consensus that you're not doing the right job, but I fundamentally believe that talking about difficult cases is best for the whole team" Director 1

8.3.4.2 Sub-Theme 4.2: Team Dynamics and Collegial Conflict

In contrast to individuals working in collaborative teams, some staff in all professional roles described consensus meetings in environments that are closed and objectionable, with a lack of respect for colleague's opinions.

"Well you might call it consensus, but I don't think I would, I hate it. …. It's alright if you work in a fair department where everybody's views are respected, but this is not the way ours works at all" Radiologist 2

"There's one Radiologist that stands there, and she doesn't really say very much, but she just harrumphs when she thinks that she might have to look at, see a patient that she doesn't agree with coming back" Consultant Radiographer 3

"Sometimes you know even in an assessment clinic you hear comments about why has this been recalled and you, you know sometimes there isn't enough respect for other people's opinion, and that's why it's easier for us here to have arbitration you know one third reader who makes that decision" Director 4

In these environments, all professional groups identified dominant personalities as an underpinning factor and collegial conflict as a critical source of dissatisfaction with consensus group review (Appendix 13 Table 47-4B). As a consequence, there were weak team interactions with a lack of consultation and discussion. This likely reflects why some survey respondents scored teamwork and group interaction low in the survey.

"We had a very dominant unit Director in breast screening whose opinion was - yes, very yeah, worth a lot more than anyone else's and basically the meeting was a bit of a sham" Radiologist 1

"It's particularly difficult we did use to have some real, real personality problems a few years ago in our department. And it got very bad indeed where there were you know, every consensus meeting we had on a particular day of the week which had a particular group of people working together was an out and out battle…. but then it got to the insulting stage where insults were thrown around ……and the trouble is the actual practice that is threatened by these sort of bad relationships within a team where there's no respect for each other" Advanced Practitioner 1

The pressure exerted by dominant individuals who are unwilling to change their view negates the process of consensus (Appendix 13 Table 47-4B). One Radiologist reported they were unable to make an independent judgement and felt compelled into making decisions they do not necessarily agree with

"The person behind me will be on my shoulder saying I wanted to bring this back because of this and I really think that, and you haven't even had a chance to look at the pictures, and so they're already getting their opinion in there, and it's really difficult to put that aside and think what would I have recalled, would I have recalled this and it's very difficult to make a non, you know make your proper decision with the rest of the group when somebody is in your ear saying I think that should come back and I'm really worried about that, when actually nobody else agrees with them, but you know it's a difficult position to be in" Radiologist 2

The tensions reported related to power politics or decision power and control within the team. Four interviewees (all professional roles) described these power battles as relatively well-entrenched and almost accepted, if begrudgingly. Furthermore, significant authority battles were described not only between different professional roles but within them (Appendix 13 Table 47-4B).

"There's too much people, sort of trying to; it's a power struggle, you know. Perhaps it's perhaps just the environment that I work in, but it is a bit of a power struggle Radiologists against Radiographers and Radiologists against Radiologists and all trying to use their little bit of power to decide what happens" Advanced Practitioner 1

Hierarchy was also mentioned as concerning in the culture of some units; all of these comments were negative and affected the voice behaviours of both Radiologists and Radiographers (Appendix 13 Table 47-4B).

"That's particularly true of film reading Radiographers. In one centre I visited as an inspector they didn't actually get to give an opinion. That wasn't their role; they were there to learn they were told" Radiologist 3

"My concern would be that in places where certain you know maybe a very dominant unit Director said oh no, we have arbitration meetings and they work very well for us thank you and other people are sitting there silently thinking no they don't. How do we hear that?" Radiologist 2

Participants also reported having their decisions undermined or overturned. Instances of being ignored or disregarded and experiences of being disrespected have resulted in some film readers being hesitant to speak up. Instead of engaging proactively in the decision-making process, they resort to quiet speech and asking colleagues to 'speak-up' on their behalf (Appendix 13 Table 47-4B). A Consultant Radiographer described how the Advanced Practitioners hope to secure her support for their decision in advance of the group discussion.

"I know the other two Radiographers they find it quite hard to make their voice heard and they can get trampled all over. They tend to hang back, or they will talk to me before we go into consensus. Oh, I've put this lady down, I really want her to come back, so if she, when we talk about her can you back me up. Well I am

thinking I won't back you up unless I actually agree with what you're saying, and so that can put me in an awkward position because they are like desperately wanting me to say, but I keep saying to them you've got to speak up for yourself, you know, this is your opinion that's going down and you need to have, find your own voice and say what you think. Consultant Radiographer 3

Interestingly, this Consultant Radiographer conveyed that she was able to defend her decision to recall, and not be overridden on a case, as she could specifically allocate the recall to her assessment clinic for workup; a facility that Advanced Practitioners do not have.

One Radiologist stated how individuals with high sensitivity and specificity (several standard deviations away from the centre on the BSIS graph) must not be intimidated by colleagues as their judgement is probably correct and they are a valuable person from whom to learn.

"I always tell people know which quadrant you're in, if you're in this quadrant let everyone know and don't give way you know hold on to what you know because you're right, so don't be cowed down by people who are saying I don't agree" Radiologist 3

Two interviewees (Director and Radiologist) expressed that their consensus review meetings were now more democratic, and this was achieved by introducing a 'golden recall alarm' or 'joker card'. These systems were implemented when an individual felt strongly that a case should be recalled, but their decision was in the

minority (Appendix 13 Table 47-4B). It was considered to support staff, avoiding the need to verbalise their justification.

"We used to have a lead, but now people can play a joker card (laughter). If one individual wants to call and no-one else does, they can use their joker card" Director 7

A more egalitarian approach in consensus meetings was also taken, but it was acknowledged that this process requires individuals to be willing to change their opinion (Appendix 13 Table 47-4B).

"I've had lots of cases where I've called it normal, and arbitration has called it back, and I put my name at review to calling it back" Locum Radiologist 1

An important issue raised was that, in order to form an independent opinion on a case, individuals require time for contemplation and a physical environment that is conducive to supporting the diagnostic process. In units that involve large group numbers, this was reported as difficult to achieve (Appendix 13 Table 47-4B).

"But before you can have a voice in a consensus meeting you've got to have time to consider the images, I find that really difficult, another reason I don't like consensus" Radiologist 3

Therefore, two interviewees questioned how effective consensus meetings are with their internal group dynamics. This was felt to reinforce recall of benign lesions in one unit; and considered 'disastrous' if the dominant voice is an over caller, particularly if a unit is trying to reduce their recall rate. The issue of cost-effectiveness, associated with the resources (time and people) to convene the meetings, was also raised.

"And it's expensive too by the time you've got eight people in a room if each one of them is going to wait for the other seven to have time to form an opinion it's going to cost you a great deal of money and a great deal of reading time to get that job done" Radiologist 3

8.3.4.3 Sub-Theme 4.3: Accountability

Analysis of survey responses identified that consensus review is being used for protected learning and education, with collaborative decision-making. Interviewees in all professional roles reinforced these concepts, as demonstrated in Table 47-4A (Appendix 13), but self-assurance was also a predominant factor (Appendix 13 Table 47 -4B). Three participants (Director, Consultant Radiographer and Advanced Practitioner) confirmed that readers are encouraged to recall cases for discussion, and subsequently, a relatively large proportion are returned to routine screening following review. Although the belief is that there is no impact on the units recall rates, it will, however, be reflected in an individual's performance report on BSIS. Hence, these individuals may have high recall rates relative to peers; this is

important as recall rates is one of the performance metrics on which a reader is judged.

"I guess it depends if your normal practise is to bring lots back to consensus like ours, is bring lots back to consensus and then chuck lots out …we probably chuck out more than half of them. So, anything we've got the slightest concern about really or that we just want to discuss with someone else so it's not like if we put it down it must come back. I think we probably review more because I think people have the confidence to put things down" Director 5

One free text comment in the survey raised the fact that a group consensus review devolved responsibility. This perception was also repeated in the interviews by three Consultant Radiographers, remarking that consensus relieves the burden from one individual and shares the accountability (Appendix 13 Table 47-4C).

"I've had the luxury of always being part of a team, so it's always being discussions with that, so I've never had that single responsibility" Consultant Radiographer 1

One Breast Clinician reported Radiologist colleagues' avoidance of arbitration because 'they don't want to make the final decision'. Staff who are undertaking a third read arbitration have ultimate responsibility, and this was deemed stressful and a higher risk to the individual. Conversely, one Radiologist considered a third read was equivalent to a group review representing 'just another opinion' but acknowledged that individuals could make wrong decisions on cases which may be challenging.

"Obviously your decision is more on one person but why that should be any different to a group thing. You are not going to get them all right, that's the point of double reading because it difficult, it's not easy is it. So, it's more a risk to the one person who's deciding, but I think it's a better system" Radiologist 2

Assumptions about legal litigation being greater in breast compared to other radiology specialities were described and highlighted by one Director. This Director considered that, from a medico-legal perspective, a group review of cases offered safer practice.

"I think it's better from a medico-legal perspective to have, if something should subsequently turn out to be a cancer, then it's better to have five people looking at that film than one. You know, so that is another consideration to think of" Director 6

The potential of litigation was also expressed as an area of concern for Radiographer third reader arbitration, with the belief that Radiographers would find a lawsuit tougher to cope with compared to medics who were considered to be more accustomed to this throughout their career. A Consultant Radiographer believed that the public did not perceive parity between allied health professionals and medics and was concerned that inadvertently returning a discrepant cancer case to routine recall is an area that would be pursued legally (Appendix 13 Table 47-4C).

"Unfortunately, breast radiology is one of the ones where it's actually; you get more claims than other areas. Doctors live with it, they know about it, it's an extremely

unpleasant situation to go through, and I'm just, I'm very personally concerned that I don't want any of my film readers to go through that situation and it's certainly not something to dismiss. It really is a very serious thing to have somebody bringing a medico Legal litigation against you" Director 6

The interviews have revealed that team dynamics are complex and a crucial factor in the success or failure of group consensus review. As discussed earlier in this chapter, external arbitration may assist with normalising consensus nationally and can provide an alternative in maintaining opinion diversity. A further option is an independent review of the discrepant films by several reporters, with the results aggregated to inform the final outcome. Potentially, this may be more time efficient as there is no requirement to convene a meeting with multiple attendees; images could be independently reviewed in such a reporting session. The potential role of AI as a second or even third reader is discussed in the following Chapter (9).

8.3.5 Theme 5 - PHE Guidance Factors

The survey responses demonstrated that only a small number of units were implementing Radiographer third reader arbitration/lead of consensus following the publication of the PHE arbitration guidance. Theme 5 builds on those responses and incorporates participants' views on the content of this guidance, recommendations for improvements and perceived barriers/facilitators to implementation. Thematic analysis revealed that these factors were either related to the guidelines themselves or were individual professional factors. Guideline factors incorporated the following themes; evidence strength and quality, and the clarity of the guidance. Individual professional factors included a lack of agreement with elements of the guidance, the

inertia of practice, appropriateness of the guidance, and the implementation climate/capacity for change**.**

8.3.5.1 Sub-Theme 5.1: Guideline Factors

Although the PHE guidance was sent to participants before the interview, two current and one former Director indicated a lack of awareness in general or unfamiliarity with the guidance criteria (Appendix 13 Table 47-5A).

"To be honest, I cannot picture what that guidance says at all" Director 1

- **Evidence Strength and Quality**

Interviewees in all professional roles had divergent views on the PHE arbitration guidance overall, some describing it as vague, nebulous, lacking evidence and detail (Appendix 13 Table 47-5A).

"The evidence seems very weak in it to me. This is very generic; this is not saying, this is not looking at your, there is nothing in there that shows you are performing to a high standard. You are just ticking the boxes; you are reading the required number of films" Consultant Radiographer 2

Conversely, others considered that it should not be too dictatorial (Appendix 13 Table 47-5A):

"I always like the guidance to be much more black and white, but they can't do that because unless you, unless you get to the point of dictating to everybody screening how they should screen read particularly" Breast Clinician 1

One Director believed that the guidance would be more useful if it specified blind reading, which was considered to make a point of difference, and there was an expression of wanting arbitration practices to be more comparable between units (Appendix 13 Table 47-5A).

"I think the guidance could be more helpful in that sense as in being sort of prescriptive about things like reading, reading blindly I think maybe they should bite the bullet and specify that" Director 5

A Director emphasised that the guidance lacked specific advice on how to select individuals for third reader arbitration or the composition of a consensus group.

"I think the arbitration guidance is a little bit; it's unclear because it says that you could choose to do consensus, but it doesn't then state whether it suggests that somebody who is an experienced film reader, but it doesn't insist there should be an experienced film reader present and it also doesn't say what if the experienced film reader who is present at consensus happens to be one of the, has already read those films so has already expressed their opinion" Director 2

However, as discussed previously, there are no evidence-based metrics upon which to measure arbitration performance. Hence, it was acknowledged that the guidance is constructed on expert opinion.

"There's always a reservation about doing that making a statement you know that's not evidence-based. Well you might say well lots of it isn't, but there's lots of

it saying this is the best advice that we can give based on what we know" Radiologist 3

- **Clarity of the Guidance**

Interviewees in all professional roles identified some ambiguity in the guidance, and subsequently variation in the interpretation, which mainly related to two statements (Appendix 13 Table 47-5A). The first statement:

Participate fully in assessment clinics including decision-making (working to Consultant Practitioner level)

"How I would interpret that guidance, that those people, those retired people if they are not doing assessment clinics now, then they are not really eligible according to those guidelines" Director 6

One Director described how, after reading the guidance, she had taken a very 'safe approach' and had stopped Advanced Practitioners and an experienced Radiologist from arbitrating. Although the Radiographers were taking part in assessment clinics, they were not responsible assessors who could make the ultimate decisions on patient management. The guidance was, therefore considered counter-productive in this unit, resulting in a loss of flexibility.

"Prior to the arbitration guidance being brought out we used to have our Radiographer film readers would arbitrate……but when the arbitration guidance

came out because it says that they must take full part in an assessment clinic including decision-making, which isn't something that our film readers do currently, so we had to stop them arbitrating. I also had to stop a very experienced Radiologist who's retired and now just does film reading, she could no longer arbitrate because she does not take part in assessment clinics" Director 2

The Director expressed the opinion that because the unit was imminently due a QA visit she therefore felt she had to implement the guidance to avoid criticism from the review team, but 'putting hurdles in the way' was restrictive and the guidance was considered 'a slightly blunt instrument'. The resulting implications of this change in practice were negative, with the possibility of arbitration cases breaching the two-week standard for results and the three-week standard for assessment.

"There's a case sitting waiting to be arbitrated; you know particularly in holiday periods. You've got experienced film readers, but they don't fulfil the criteria to be an arbitrator so that case just has to breach because there's nobody here that can arbitrate it" Director 2

A Radiologist involved in the consultation process for the guidance clarified that initial discussions had included whether Consultant Radiographers should be allowed to arbitrate, which was deemed 'radical'.

"It was pretty much earlier on agreed, not by me actually, but that Advanced Practitioners probably weren't the right people to arbitrate, and that was partly

because they thought it might be tricky politics in units etcetera and so it seems a radical departure to actually be adding in Consultant Radiographers to the mix" Radiologist 3

The same member of the consultation group raised the point that some Advanced Practitioners were undertaking Consultant level duties but did not have the Consultant title.

"Some people didn't get the title Consultant but were still doing all the Consultant activities like attending assessment etc. that we should say that if they're at the level of a consultant practitioner that's ok and leave it to the discretion of the Director of screening, but that was the intention and the background to why it says what it says" Radiologist 3

From the researcher's perspective, this implies a lack of understanding about Consultant and Advanced Practitioner roles as the Consultant Radiographer role is not purely based on autonomous expert clinical practice but incorporates three other domains (professional leadership, service development; research and evaluation and education and professional development).

The second statement considered ambiguous by the interviewees was:

The Society and College of Radiographers (SCoR) can provide accreditation of Advanced and Consultant Practitioners regarding the four pillars of practice which

include: leadership, CPD and education, clinical practice and audit/research capabilities

Seven interviewees (all professional roles) commented that they did not understand the context of the accreditation reference within the document or that the information was contradictory regarding advanced practice (Appendix 13 Table 47-5A).

"There's just this weird statement about the society can provide accreditation. I don't know it seems out of place that sentence hmm because it sort of specifically says accreditation of Advanced and Consultant Practitioners but then above it's fairly specific that only referring to Consultant Practitioners, and Advanced Practitioners are not, yes it seems a little bit out of place, doesn't it? Why does it say, because you could be an accredited Advanced Practitioner, but then you still can't arbitrate it seems a very out of place statement, doesn't it?" Director 5

The value of Radiographer accreditation with relevance to arbitration was therefore considered by these interviewees as superfluous and of no tangible benefit, as Trusts indemnify individuals for undertaking the extended scope of practice (Appendix 13 Table 47-5A).

"I personally wouldn't want to go through accreditation. Ermm I am happy with what I am doing, and the trust are happy with what I am doing I don't feel that I

need, and I've got a masters. I don't feel like I need to put myself through all of that box-ticking" Consultant Radiographer 3

8.3.5.2 Sub-Theme 5.2: Individual Professional Factors

- **Lack of Agreement**

Six interviewees identified several attitude-related barriers. There was a lack of agreement with more guidance in general, with the belief of too much bureaucracy already.

"We're just trying to make it too complicated there are so many protocols already regarding breast and just having another one for arbitration I think it would just be a negative impact really" Director 4

There was also a lack of agreement with specific aspects of the guidance relating to professional factors, which included working at a Consultant level, accreditation, and the perceived lack of applicability to individual clinical situations. The restriction of staff working at a 'Consultant level' was criticised and deemed by two Radiologists to be a negative and offensive statement (Appendix 13 Table 47-5B).

"I think that's what pisses me off about kind of very prescriptive things from NBSS saying you can't have them (Advanced Practitioners) do it you must have them do it. So, I think it's a bit insulting really to the Advanced Practitioners who are experienced. Well, then that would mean that our Advanced Practitioner couldn't. I think then you're limiting yourself to a handful of Radiographer Consultants

throughout the country. Well, I think that statement is unhelpful, I would say. So, if you had a retired Consultant who is screen reading but never attended you know the MDT, I would have a problem with that, and they probably shouldn't third read" Radiologist 1

Advanced and Consultant Practitioner accreditation is currently voluntary, and therefore one Radiologist considered this should not sit within national guidance.

"So they need to decide is accreditation desirable, essential, if it is then everybody has to do it they're on a list and then you know there is criteria for being on that list and then you know so if it's ad hoc at the moment being on that list you can't then put that into guidance. It doesn't seem to be relevant." Radiologist 1

A Radiologist involved in the consultation document clarified that the purpose was to direct people to the society's website and how accreditation can be achieved.

"So, it's stuck in there to make sure those people are saying well you are talking about ermm Consultant Practitioners, but they can't become those because that's difficult there is a link to how you get to be one" Radiologist 3

- **Inertia of Practice**

A further perceived attitude-related barrier was the inertia of practice (Appendix 13 Table 47-5B). An Advanced Practitioner reported difficulties in changing professionals' reporting habits and procedures, and that implementation of a change only occurs when there is a staffing crisis. Apathy to implementing change

was professed to be a protectiveness of professional roles. Conversely, one Radiologist considered that the organisational culture, rather than professionals, influences the willingness to change.

"It's likely that we are dealing with cultural stuff that's now stuck and I think that's what we've got. I think when something is completely new it's easier to fix it when people have been doing it one way for a while some other way its rather more challenging" Radiologist 3

- **Appropriateness**

There was also a lack of agreement with the guidance regarding its applicability and timeliness (Appendix 13 Table 47-5B). One Radiologist reflected that local arrangements had developed in advance of the guidance because of the pressures on services, or third reader arbitration was automatically assumed when Radiographers qualified as a film reader. In this unit, there was no minimum period or professional role stipulated before undertaking the function.

"No, it didn't have any effect on us at all, we just carried on doing what we were doing. Ermm it was just always like that right from the word go as soon as (name redacted) she was our first Radiographer film reader, and it was just taken as read that she would do it, I don't even think there was a discussion, it just happened so when the other girl qualified she started to do it and as soon as I qualified I was doing it as well. So, we didn't have the like you have to be doing it for two years before you can actually arbitrate as a third, we just did it as soon as we qualified" Consultant Radiographer 3

Additionally, a constraint on applicability was expressed because interviewees thought that the guidance was explicitly tailored to third reader arbitration and not encompassing of the characteristics and variance in a unit's practice (Appendix 13 Table 47-5B).

"The group that's written this will they know there is such variation in practice they seem to have written for a very specific situation where there's a single independent third reader arbitrating; perhaps they didn't appreciate there's such a huge range, huge difference in practices and to a lot of units such as us it wasn't helpful or it didn't really mean much as it didn't relate at all to what we were already doing" Director 5

- **Implementation Climate/Capacity for Change**

In three units, the guidance had supported the implementation of Radiographers as sole third reader arbitrators or leads of consensus review meetings. In accordance with the survey responses, the key driver for implementation was a service need due to the lack of Radiologists and the inability to recruit to vacant posts (Appendix 13 Table 47-5B).

"Because we're short-staffed" Advanced Practitioner 1

Delegation had increased the workforce able to undertake the task, with the main benefit that cases requiring review were now dealt with promptly, avoiding delays and potential breaches (Appendix 13 Table 47-5B).

"With the new practice, it just meant that we weren't have to wait because sometimes the Radiologists aren't around on a Friday. So, if they weren't around you couldn't do it" Consultant Radiographer 1

In another unit, Radiologists primarily undertook the third read and Radiographer arbitration was only utilised as a necessity.

"We have one Consultant Radiographer who can, who does do arbitration but not very often and it's only really like today she did it because the other two people that were there had read the films, so she had to do it" Radiologist 2

The interviews revealed that, in reality, some units are using semi-retired Radiologists (no longer actively undertaking assessments) and Advanced Practitioners to undertake arbitration, which illustrates a significant shortage of qualified personnel in the breast screening programme.

"People are being dragged in from retirement, and people who wouldn't normally be involved in the process are being involved in the process" Director 6

The interviews indicate a need for leeway in not working exclusively according to guidance, as flexibility is required in the current staffing climate (Chapter 1 & 6). As indicated, individual units are using a combination of what works, and what is needed, based on the staff they have. One Radiologist explained that when she was Director of the unit, the visiting QA team raised the issue of the unit utilising Advanced Practitioners as arbitrators. However, this is a small unit, and discontinuing this would have resulted in failed NHSBSP standards.

"QA visits did question us on it because I was unit Director at that point and I was like do you know what we haven't got a hope of making any targets if we stop that because if you think about it almost every week in the year one of us is away study-leave annual leave you know four people so that's like 25% of our screening workforce gone and there wouldn't be somebody to do the 3rd reads if we relied on a Consultant only… we've got what we've got you know; we wouldn't be able to change" Radiologist 1

As discussed in section 8.3.3.2 of this chapter, it may be that in time the review of film reader statistics via BSIS will be able to demonstrate individuals who are consistently not only sensitive but specific in film reading, and Trusts will continue to indemnify varying professional roles as arbitrators. Overall, barriers to Radiographer arbitration were identified by participants at both an individual and organisational levels. It is acknowledged that these results are from a cross-section of time and that they may differ if repeated. An overview is provided in Table 48.

Table 48. An Overview of the Perceived Barriers to Using the PHE Arbitration Guidance

1. Knowledge-related barriers - Unfamiliarity with the guidance - Lack of knowledge of content (recommendations for delegation)
2. Attitude-related barriers Lack of agreement with the notion of more guidance – too much bureaucracy Lack of agreement with this specific guidance: - - Appropriateness – lack of applicability in some units, not encompassing of unit variance and characteristics. Timeliness - produced too late, units have implemented local strategies. - Evidence strength and Quality –perceiving them as too vague, lacking detail or too prescriptive to apply, not practical and limiting flexibility - Lack of clarity - ambiguity in the interpretation of i) *Participate fully in assessment clinics including decision-making (working to Consultant Practitioner level)*, ii) *SCoR accreditation* - Lack of self-efficacy (Radiographers) – fear of litigation - Lack of outcome expectancy – increased recall rates without an associated increase in CDR
3. Organisational barriers - Organisational constraints – policies on reporting practices, inertia of practice - difficulties with changing habits and procedures

8.4 Summary of Findings

The interviews show that historical, cultural elements dictate the current variation in reporting and arbitration practices. In some units, in-house training and non-blind reading/arbitration are resulting in conformity of practice and potentially biased decision-making. There is a potential need for change with some clinical trials mandating blind reading, but the ideology of becoming paperless was contrasted with the practical difficulties and safety aspect of ensuring the right results process. Although many interviewees theoretically supported a complete electronic reporting system as the way forward to support blind reading/arbitration and ease of audit, the practical process of actually setting this up in terms of IT infrastructure (NBSS) was seen as far more challenging.

To date, little consideration has been given to how the arbitration process may be quality assured because of the statistical difficulties. BSIS performance data may help inform which individuals within a unit are best suited to undertaking solitary third reads or lead consensus group review. For some smaller units, the concept of external arbitration provides an alternative option. Although many participants could see the potential benefit to this, those who defended maintaining current practice highlighted a potential delay in results.

Personalities emerge as a dominant factor, not only in consensus group review but also in the decision-making aptitude of Radiographers, potentially preventing delegation of third reader arbitration. A further barrier to delegation is performance-related (recall rates), which was deemed to be training-related.

The next and final chapter aims to synthesis the results of the different elements of the study. This is achieved via triangulation of the findings from the literature review, survey data, KC62 performance data. The chapter will also highlight implications for practice and subsequent recommendations and suggest avenues for further work.

Chapter 9. Integrating Evaluation, Discussion, Conclusions and Recommendations

This thesis aimed to develop a broad and in-depth understanding of reporting and arbitration practices within breast screening in England and to investigate to what extent specific systems have worked better in differing units. This final chapter attempts to triangulate the research findings described in earlier chapters, highlighting methodological issues and exploring the strengths and limitations of the research. Consideration is then given to the implications of the study findings for policy and practice and how these might influence future service provision, followed by consideration of the potential impact of future advances in technology on breast screen reporting and arbitration. The chapter concludes with suggestions for future research.

The study used methodological triangulation (as described in Chapter 5.3) to integrate data from the literature review, survey data, KC62 data and interviews to corroborate observations, highlighting any similarities and disparities in findings.

Section 1: Triangulation of the Research Findings

Several interacting issues were identified. The first key issue of service variation emerged from the integration of both quantitative and qualitative data demonstrating variance in all elements of reporting and arbitration practices as well as recall and cancer detection rates.

9.1 Issue 1. Service Variation

- **Reporting Practices**

Quantitative data (survey responses Chapter 6.9.4.2) reveals that although double Radiographer reporting is the norm in some units, described by interviewees as an evolution of practice, in other units reporting restrictions are present. The pairing of film readers is predominantly based on professional role, with one reader being a physician or Radiologist. Only a few respondent units paired on reader performance measures (recall rates and cancer detection rates) as advised in the NHSBSP Quality Assurance (QA) guidelines (Public Health England and PHE 2011).

Performance of individual reporters has been stated as the leading cause of variation in the accuracy of screening mammography (Miglioretti et al. 2007, Elmore et al. 2009, Skaane et al. 2008, Duijm et al. 2009). Age, years of experience, and the number of yearly mammograms read are classed as contributory factors. Variation in the age of the reporter and years of experience is inevitable, but a central finding highlighted by interviewees in this study is that age and years of experience do not necessarily correlate to high performance. Scott and Gale (2007) have suggested that years of experience and reporting volume determine sensitivity, rather than specificity measures. For arbitration, specificity is required, and interestingly Scott and Gale (2007) identify that reporters with 1-5 years' and 6-10 years' experience performed significantly better in terms of specificity than those with 11 years and more experience. However, this result was based on a test set of cases and not real-life performance.

Recommended annual reading volumes vary internationally. In the US, single reading or single reading with CAD remains conventional practice, with a requirement to read a minimum of 480 mammograms (screening and diagnostic) per year (960 in 2 years) (Mammography Quality Standards Act Regulations; FDA 2019). High annual reading volumes in the US and Canada are defined as 2,000-3,000. Several European programmes advocate an annual minimum of 5,000 mammograms per reporter to maintain high performance (Perry et al. 2007, Public Health England 2011, European Commission Initiative on Breast Cancer 2019). However, these figures are based on independent double reading with consensus, but this study demonstrates that the majority of survey respondents (63% -31 units) read non blinded, and this may, therefore, influence the association between reading volume and reading performance. There is a requirement in the NHSBSP to first read 1,500 films which then reflects an accurate profile of an individual reporter. In this study, the number of years of reading of the varying professional roles was recorded, but not the volume of reporting.

Generally, observational studies (using non-blinded reading) have not demonstrated a clinically relevant association between the volume of films read and sensitivity nor CDR (Buist et al. 2011, Théberge et al. 2014, Duncan and Scott 2011, Cornford et al. 2011). However, although the association between reading volume and specificity is inconsistent (Buist et al. 2011, Théberge et al. 2014, Duncan and Scott 2011, Cornford et al. 2011, Alberdi et al. 2011) the majority state a higher specificity with more films read (Smith-Bindman et al. 2003, Théberge et al. 2014, Alberdi et al. 2011). Hoff et al. (2019) demonstrated a decrease in sensitivity when annual reading

volumes exceed 10,000 and Cornford et al. (2011), suggested an upper limit of 25,000 over three years. This differs from the Duncan and Scott (2011) study that did not support a drop in reader performance at this volume of reading. Hoff et al. (2019) is the only study to date analysing the effect of reading volume on performance with blind reading of digital mammograms and consensus. Their study suggests optimum performance could be attained with annual reading volumes of 4000–10,000 mammograms.

An important finding from Hoff et al. (2019) was that more cases were dismissed at a consensus that had been recalled by low-volume readers compared to high-volume readers. Consequently, they conclude that units with low-volume readers may have more arbitration cases, but this is based on independent (blinded) reading. Hoff et al. (2019) also endorse that for consensus to improve consistency and performance (lower FPR), at least one high volume reader (Radiologist) is present. These findings support the PHE arbitration guidance, which states that staff undertaking arbitrations should read >5000 films per year mammograms.

With the increased reader profile information now obtainable from BSIS, if units have the capacity of pairing, it may be more effectual to base pairing on performance parameters, and volumes of films read rather than job roles. With a decreasing number of staff available to report, optimal reading volumes may become of more significant concern in the future.

- **Arbitration practices**

Capacity is an issue affecting the scheduling of arbitration review, which is often ad-hoc, time- pressured and governed by workload pressures and logistics. Integration of both types of data (survey data chapter 6.9.5.2 and qualitative data chapter 8.3.1.3) demonstrate that units swap processes (from the third reader to group consensus or vice versa) based on a desire to balance performance statistics with staffing levels and resources. Both processes are known to reduce recall rates, thereby improving the specificity of a programme. In the literature, 39–50% relative reductions are reported with consensus recall policies (Brown, Bryan, and Warren 1996, Anttinen et al. 1993, Hofvind et al. 2009, Dinnes et al. 2001), and 25-32% reductions with arbitration (Ciatto et al. 2005, Duijm et al. 2004). These figures would be in keeping with interviewees stating that about 50% of cases are returned to routine recall at consensus review. Analysis of KC62 data showed that there is no statistically significant difference in mean overall recall rates dependent upon the arbitration type, or cases arbitrated (discordant and/or concordant) for both prevalent and incident screens. These results are in keeping with past research (Blanks, Wallis, and Moss 1998).

Quorum requirements for consensus and the decision-making process (majority, experience weighted, profession weighted) are also variable. Interview data corroborated that the lack of standardisation is a result of historical-cultural practices, with organisational or professional culture defining departmental processes, policy, norms, and tasks. Over time, units develop a specific culture, with attitudes, standpoints and work methodologies, which contribute to the

establishment and consolidation of these identities and cultures (Diamond and Allcorn 2009).

- **Recall Rates and Cancer Detection Rates**

Recall rates are one of the main performance parameters used to determine the overall accuracy of the breast screening programme (Gur et al. 2004, Yankaskas et al. 2001a, Otten et al. 2005, Smith-Bindman et al. 2003). The literature review undertaken highlighted that there are difficulties in establishing an optimal recall rate as there is a balance between the benefit of detecting a small number of additional cancers relative to the increased number of false-positive recalls (Schell et al. 2007, Mohd Norsuddin et al. 2015). International variance in recall rates and related performance measures (Elmore et al. 2003, Roman et al. 2013, Hofvind, S et al. 2012), is commonly explained by differences in the screening interval, reporting procedures, differing recall policies, and legal significances. However, within the NHSBSP, all units operate according to the same guidance, and therefore, the variances should be minimised. In the current study, performance metrics were analysed to map consistency across units. Analysis of the KC62 four-year data demonstrated significant differences between the 80 screening units for both SDR and recall rates. Average overall recall rates ranged from 2.14% to 6.92%, but in particular, there was a more significant variance in prevalent recall rates ranging from 4.2% to 13.7%. The reasons for this remain unclear. Although units with a low PPV and a low SDR may be considered a poor performer, there are multiple contributory factors, not all of which are in the unit's control (Bennett and Blanks 2007). A degree of variation is due to differences in the population being screened

(population change, age distribution, private screening concurrent with NHS screening) (Carney et al. 2003, Kavanagh et al. 2000). Further differences in the population screened may include the use of HRT, presence of symptoms and breast density. As discussed in Chapter 2 (2.3.4) breast density affects mammography performance, (Carney et al. 2003) and should, therefore, be considered when comparing performance measures, but it is not currently a requirement within the NHSBSP to record breast density.

The conceptual diagram (Chapter 2, Figure 3) demonstrated the complexity of factors that can affect recall rates. In the literature review, several studies (Elmore et al. 2003, Yankaskas et al. 2001a, Otten et al. 2005, Schell et al. 2007) reported that increases in recall rates were not directly associated with improved CDR. In contrast to this was the USA study by Grabler and colleagues (2017) which reported that a recall range of 12%- 14% would provide optimal cancer detection rates. Analysis of KC62 data in the current study found no statistically significant correlation between the 4-year average recall rates and small (<15mm) cancer detection rates (prevalent and incident). However, there was a statistically significant, weak/moderate positive correlation between the prevalent/incident recall rates respectively and SDR. The peak incident SDR in the data occurred with a recall rate of 3.781%. Although the location of the cut point should not be over-interpreted, the data in Graph 22 (Chapter 7.7.4) demonstrated that the sensitivity did not increase beyond this rate.

Notwithstanding its limitations, this study's results concur with the recent analysis of more than 11.3 million breast screening exams from the English Breast Screening

Programme which concluded that there is an optimum range for recall (Prevalent 4.6% to 7%, Incident 2.6% to 4%) to maximise the detection of life-threatening cancers and minimise the harm from excess false positives (Blanks et al. 2019). The results presented in this thesis emphasise the potential to improve effectiveness in some units by reducing recalls and false positives while maintaining acceptable CDR. However, Burnside et al. (2018) propose that a minimum recall rate should be implemented for the NHSBSP as the authors established a statistically significant negative association of recall rates and interval cancers; if too low the benefits of screening are diminished.

The four-year average SDR rates analysed in the present study demonstrated a 2.26-fold variation (0.93 – 2.1) for prevalent SDR with a smaller 1.47-fold variation for incident SDR (1.21-1.78). However, when the 80 units were ranked in ascending order, there was considerable variation in a unit's position with data influenced by extreme values in single years. Conversely, consistent patterns in recall rates whether high or low were demonstrated over the four years, and this therefore likely reflects the cultural norms developed in reading and arbitration practices in the individual units as described by the interviewees. However, analysis of the arbitration strategy demonstrated no statistically significant difference in unit recall rates for <15mm CDR and SDR for both prevalent and incident screens. This was irrespective of programme size. This differs from the study of Blanks et al. (2002) which stated a lower SDR for the small programmes and suggested that performance (PPV and CDR) was marginally poorer compared to medium or large-sized units.

- **Double Reading/Blind Reading**

There is convincing evidence that double reading increases sensitivity and CDR, with the literature reporting 9- 25% of cancers detected by only one reader (Euler-Chelpin et al. 2018, Taylor-Phillips et al. 2018, Liston and Dall 2003). Within the present study, two Directors stated that single reader cancer rates were about 25% in their unit. However, since the introduction of digital mammography, the value of double reporting is often questioned in terms of the cost, infrastructure and time implications (Posso et al. 2016). As single reader cancers tend to be smaller and less advanced (Taylor-Phillips et al. 2018, Dinnes et al. 2001) any extra expenditure should be offset against the costs of treating late-stage disease (Dinnes et al. 2001, Duijm et al. 2004). Conversely, opponents to breast screening would suggest the non-invasive and small, low-grade disease represent overdiagnosis. Although all units in the NHSBSP currently double read, it is evident from this research study that the variation in reading practices and culture of units greatly influence the number of cases recalled to assessment.

Survey results (chapter 6.9.4.1) revealed that the majority of respondent units read non-blinded, which again was stated to be historical practice. It is evident that some interviewees perceive a change to fully blind reading as 'a bit of a culture shock', but if introduced this may negate the requirement for readers (in some units) with higher recall rates to be limited to the first reading. FRQA data for second reads would also then be useful, as they are currently less reliable because of the multiple variables described (See Chapter 2). A predominant theme identified in the interviews (Chapter 8.3.1.2) was the concept of following on; which was deemed

especially applicable to newly qualified film readers. In theory, if readers are not unduly influenced, their recall rate should be similar for first and second reads. Analysis via BSIS may inform if the second reader's performance would be inferior with blind reading, as suggested by one Director (in the current study) and would provide a more accurate profile of an individual's reading performance.

It is unclear whether complete blind reading would expose more significant differences in reader ability. Analysis of quantitative data (KC62) has demonstrated no difference in mean recall rates between units based on reading type (blinded vs non-blinded), but it is unknown how many cases are subsequently dismissed (returned to routine recall) at the review. However, with fully blind reading, there are several essential issues to consider. Firstly, there is limited literature in this field. Klompenhouwer et al. (2015a) stated that, in the Netherlands, blind reading almost doubled the proportion of discrepant cases. In this setting, there was an associated significant higher recall rate, but there was no arbitration, all cases with discrepant readings were referred for further assessment. Qualitative data in this study confirm that there was a significant increase in the number of arbitration cases when blind reading was implemented. This supports the theory of a biased perspective when not blind reading. However, of more considerable significance is that Klompenhouwer et al. (2015) reported a higher, although not statistically significant difference, in CDR (7.4 versus 6.5), but also that less invasive interval cancers were detected with blind double reading compared to non-blinded reading. Interval cancers (Fong et al. 2012, Bellio et al. 2017) have worse survival than screen-

detected cancers; therefore, Klompenhouwer et al. (2015) advocated the use of blind reading.

Overall, the results from this study indicate that perceptions of the optimal method of reading (blind vs non-blind) appear to be heavily dependent upon staff personalities and whether there are dominant individuals within a team who influence others. With the majority of respondent units reporting they do not undertake blind reading; consideration should be given to standardising reading practices nationally to fully blind reading if this has the potential to improve the sensitivity of a programme.

With the expansion of the NHSBSP (age extension trial), it is imperative to maximise the effectiveness of both the film reading and arbitration process. The consequences of such variation in practice can result in a higher number of false positives, which have significant implications from an economic perspective, and adverse psychological effects upon the women recalled (Hofvind, Solveig et al. 2012, Brodersen and Siersma 2013, Kopans, Smith, and Duffy 2011). Discordant results, and how they are handled, is important. A recent study by Houssami et al. (2017) reported that discordant recalls were confirmed to be malignant in 10% of women, and Hofvind et al. (2009) reported interval cancers were higher in the cohort of discordant reads that were not recalled. Taylor-Phillips et al. (2018) also reported higher CDR on arbitrated cases that were returned to a routine screening on the previous round. Some of these cases would have been true interval cancers and others, cancers which were missed by arbitration, giving a range of 8.9%-10.3%. However, the authors assumed independent second reading and the present

research study demonstrates that blind reading and blind arbitration does not occur in many units.

While there is a recent focus on improving reader performance via BSIS feedback, little is currently done to QA the arbitration process. The difficulties associated with this were described in Chapter 8 (3.3.2) and corroborated by qualitative data. As a solitary third reader, variation in unit performance will depend upon the reporter's predisposition to detecting every cancer (reflected in their sensitivity) and acceptance of false-positive recalls (reflected in their specificity /PPV). However, information obtained via BSIS may facilitate performance-based selection of individuals to undertake the third read or the optimal composition of groups for consensus review, but another level of re-organisation may be difficult in busy environments.

9.2 Issue 2. Decision-Making and Radiographer Self-efficacy

Two key issues identified in the present study by integrating quantitative and qualitative data were decision-making and self-efficacy.

- **Decision-Making**

Third reader arbitration or leading a consensus group review requires decisive decision-making skills. Free text survey comments (chapter 6.9.4) and qualitative interview data (chapter 8 sub-theme 3.4) identify that some Radiographers might be indecisive and display more cautious decision-making than Radiologists and Breast Clinicians. The literature regarding ‘confidence' in advanced and Consultant

Radiographers emphasises that personalities are concomitant with assertiveness. Price and Edwards (2008) assert that competence and confidence are entwined personal qualities requiring a double stratagem for development. The authors termed self-belief and confidence as ‘soft' skills and, although they recognise that individuals must inherently possess these attributes, they suggest these skills could be improved with assertiveness training and unified organisational support. They also suggest that competence is improved with feedback, review of performance analysis and leadership training. As discussed in Chapter 2 (2.5.4), NHSBSP film readers are provided with performance statistics and opportunities to review interval cancer, allowing evaluation of any cancers that were dismissed at arbitration. The findings would seem to indicate that other elements such as assertiveness, leadership and organisational support may be lacking. However, this study did highlight that variable practice exists in terms of feedback and survey respondents gave very low scores for having mechanisms in place to monitor the outcome of consensus. The practice of not reviewing the routine recall cases may reinforce a pattern of recalling the same type of false-positive cases. Hence, feedback is essential for advising individuals and teams about their diagnostic performance (diagnostic error) and can benefit clinicians in modifying overconfidence and reducing the chance of repeated errors (Croskerry and Nimmo 2011, Schiff et al. 2009). In units where there is a limited group discussion of cases, separate educational sessions with a review of arbitration cases with known outcomes may prove beneficial, especially if there is a dominant leader who is a high or low recaller. Qualitative data confirms time constraints and workload pressures as factors that prevent team review.

Yeung and Summerfield (2012), in discussing expertise and self-efficacy, describe a ‘first-order confidence' which is associated with making a decision. In terms of screen reading, this relates to determining that an abnormality is present. The authors also describe a ‘second-order confidence' which refers to the probability of the decision causing favourable (correct) or unfavourable (incorrect) outcomes. This second-order confidence requires a breast screen reporter to differentiate the benign from the potentially malignant and is particularly relevant to arbitration cases where the aim is to increase the specificity. In contrast, Boldt et al. (2019) maintain that confidence is not necessarily associated with favourable outcomes, but the ability to make a decision and deal with the consequences whatever the outcome. This is in keeping with a barrier identified in the present study; the fear of liability and litigation linked to Radiographer arbitration. This is a common finding in upskilling in other AHP areas when the task-shifting involves more complexity (Colvin et al. 2013, Bhutta, Lassi, and Mansoor 2011). According to McMurray (1992), educational factors, personal factors, and experience affect expertise, and the author goes on to state that experts may be individuals who maintain a sense of perspective about themselves. To overcome the barrier of a lack of efficacy, a period of mentorship was recommended by two interviewees as a potential solution. In conjunction with the review of performance data from BSIS, this may give confidence to individuals to perform arbitration.

The literature review in Chapter 3 (3.4) also highlighted that, although humans consider their decision-making as logical, the reality is that a multitude of biases can distort an individual's perception resulting in irrational decision-making (Blumenthal-

Barby and Krieger 2015, Kahneman and Klein 2009, Elstein, Schwartz, and Schwarz 2002). These perceived misinterpretations can result in an overestimation or underestimation of the disease, which has significant consequences for diagnostic decision-making. Chapter 2 (2.7) discussed interpretative errors when an abnormality is identified but misinterpreted (Wadhwa, Sullivan, and Gonyo 2016). This applies explicitly to arbitration cases as the potential area of concern is evident to the third person arbitrator. Collaborative team-based diagnosis is advocated to avoid personal blind spots and biases (Balogh et al. 2015). Although a team-based diagnosis may be considered superior as a means of reducing diagnostic error, there is a lack of evidence to demonstrate an associated improvement in diagnostic accuracy in a clinical setting (Balogh et al. 2015). The present study did not aim to investigate the interval or missed cancer rates.

The concept of collective intelligence was discussed in Chapter 4 (4.5.1), with studies from non-medical domains (business, politics and economics) reporting an increased prominence of situations in which CI is valuable (Arrow et al. 2008, Koriat 2012, Woolley et al. 2010) with groups (individuals but acting independently) stated to outperform single individuals (Balogh et al. 2015). However, diagnostic accuracy studies demonstrate conflicting results (Wolf et al. 2015, Kurvers et al. 2015, Hautz et al. 2015, Kattan et al. 2016, Kee, Owen, and Leathem 2004, Christensen et al. 2000). Nevertheless, the use of algorithmic CI rules could be advantageous in specific breast screening units. If staffing levels preclude consensus meetings, individuals could individually report the arbitration cases during their regular reading session, with the results aggregated into an outcome. Also, this system

could prevent the sociological biases (groupthink and conflict evasion) as discussed in Chapter 4 (4.5.1), where individuals refrain from constructively expressing viewpoints (Lighthall and Vazquez-Guillamet 2015), the result of which can be a conformity of practice, as confirmed by interviewees in the current study (Chapter 8.3.1.2). A significant novel finding in the current study was interviewees acknowledging bias of their judgement based on knowing who has recalled the case. This was not only when second reading but also when undertaking arbitration and group consensus review. In this study, individuals describe learning to recall like peers, modifying their recall behaviour, which in turn can lead to mediocrity. However, further research would be required to evaluate this method in a real-world clinical environment and to develop optimal rules for combining opinions to reach an unbiased decision.

Mahmoodi et al. (2015) also describe an equality bias in group decision-making. Individuals have differing levels of competence and, although theory indicates that the opinion of individual members should be weighted by reliability (Bovens and Hartmann 2003, Owen 1989), empirical research acknowledges that this can be challenging. Markers of reliability are influenced by numerous characteristics, which include confidence, personality (Campbell, Goodie, and Foster 2004), gender (Hannagan and Larimer 2010) and culture (Mann et al. 1998). To reach a sound decision, it may be necessary to discard the view of a less proficient reader which can be challenging. Equality biases are also stated to be more problematic with small groups (Mahmoodi et al. 2015). Considering every one of equal ability and thus allocating an equal weight to each opinion may decrease the accuracy of the group

decision (Mahmoodi et al. 2015). However, Koriat (2012) also describes the problem of subjective confidence, which assesses the unanimity of the group rather than the accuracy of the decision. In situations when the majority of the group are in error, decisions dominated by the most confident individual yielded worse decisions than those of the best individual.

The literature review also highlighted that individuals could grossly misjudge their own level of competence, and their level in comparison to peers (Gigerenzer, Hoffrage, and Kleinbölting 1991, Soll 1985). To what extent these individual differences in competence are accounted for in breast consensus decision-making is unknown. As discussed previously, BSIS data will provide information on a unit's most sensitive and specific readers and this could inform the optimal group composition for group arbitration, which could take into consideration differences in ability. This raises several questions for future research. Would a more formalised approach to consensus group composition result in improved unit performance metrics? Could algorithmic CI rules improve the collective intelligence of the consensus group? The risk of conformity could be avoided by anonymity, which would require changes to the NBSS system. However, would the anonymity of the recaller improve the outcomes of consensus groups?

9.3 Issue 3: Culture and Implementation Climate

The integration of qualitative data from the surveys and interviews (Chapter 6 6.9.10, 6.10.5.1 and Chapter 8 sub-themes 1.2, 4.2, 5.2) demonstrated that a key issue was the culture and implementation climate of units.

- **Task Shifting**

Chapter 1 (1.1) highlighted the current diminishing Radiology workforce in breast screening. A fundamental approach historically utilised to address chronic shortages in the health workforce is the redistribution of tasks between different professions (Colvin et al. 2013). Task shifting is evident in other areas of AHP practice, intending to optimise the effectiveness and efficiencies of skill mix (Colvin et al. 2013).

The PHE arbitration document provides guidance for the delegation of arbitration to individuals who may act as a single arbitrator (third reader) or in consensus act as the coordinator/lead. Integration of quantitative data (Table 30 Chapter 6.9.10) and interview data emphasise that organisational culture and difficulties with changing habits are the main factors inhibiting implementation, which concurs with the theory of planned behaviour proposed by Catchpole (2013). The culture of an organisation is founded on behavioural practices that collectively result in the 'way things get done around here' (Senior and Swailes 2016). Many barriers are entrenched, as a result of historical methods of training and a change in professional identities, that have added to conflicting approaches to communication and hierarchical relationships. These findings are consistent with those identified in a systematic review of task-shifting in midwifery (Colvin et al. 2013). Attempting to influence change within the multi-faceted, and varied dynamics that exist in healthcare teams is a more extensive undertaking (Braithwaite 2018).

Cabana et al. (2001) characterised three main groups of barriers to guideline adherence: 1. Knowledge-related barriers, 2. Attitude-related barriers, and 3.

External barriers. In the current study, both data types identified knowledge-related and attitude-related barriers, but the third category related to organisational constraints rather than external barriers.

Quantitative data suggest that PHE guidance has had minimal impact on introducing Radiographer as third reader arbitrators or lead of consensus. However, qualitative data reveals valid reasons for non-delegation in specific units. Several requirements in the guidance are not currently met by Radiographers (Advanced Practitioners and Consultant Radiographers), such as autonomous decision-making in assessment clinics and actively participating in decision-making and subsequent patient management within MDT'S. In several units, the PHE guidance was considered restrictive and counterproductive, demonstrating that in reality, it may not be appropriate for every clinical setting.

It may be that responses would have been different from non-respondent units, and therefore, this may not represent an accurate national picture. Although there are only a small number of units utilising Radiographers to undertake third reader arbitration, quantitative analysis (KC62 data) showed no statistically significant difference in four-year average overall recall rates dependent upon the professional role (Chapter 7.10). This supports the statement in the PHE arbitration guidance that the skills required (decision-making with excellent specificity) are not necessarily associated with the profession of the arbitrator.

The AGREE Collaboration (2003) (Appraisal of Guidelines, REsearch and Evaluation) assert that guidelines should explain how new evidence will be monitored and

recommendations updated, if required. However, this prerequisite is infrequently performed (AGREE Collaboration 2003). To date, there has been no published follow up to determine the impact of the PHE NHSBSP guidance.

- **Collaborative Teamwork**

Collaborative teamwork is intrinsically a complex phenomenon, involving multiple factors (group size and diversity, professional roles, decision-making rules, group dynamics) which will determine the effectiveness of group decision-making. The benefits of consensus are reported to extend beyond improved diagnostic effectiveness, offering an educational process (Pow, Mello-Thoms, and Brennan 2016). This concurs with qualitative and quantitative data (team dynamics table) in this study. However, this is not a unanimous viewpoint. Integration of both data types reveals that teams exhibit a combination of positive influences (e.g. trust, openness, respect) and negative/constraining influences (e.g. disrespect, a rigid culture prohibiting open discussion). The predominant factor emphasised is that of the personalities and attitudes of individuals. Despite group consensus being the approach adopted by the majority of units, there was nothing identified by the literature review that related to team dynamics within a breast screening setting.

The team dynamics model produced by Song et al. (2015) emphasises the principal factors that support teamwork are accountability, communication and conflict resolution. Quantitative results reflect polar scoring in these factors by varying professional roles. Although these findings demonstrate an overall positive level of team dynamics, the statement ‘considering all points of view' before deciding on the

outcome was scored at the two extremes by some individuals. In the current study, this was confirmed in the qualitative data with a pressure to conform mentioned, and a high dependency on decisions from senior staff members. 69.4% of respondent units use a majority vote as the primary decision-making method (Chapter 6.9.9). Bang and Frith (2017) report that utilising a system that creates a majority and a minority vote can cultivate conflict and distrust. The qualitative data also corroborated this with several interviewees describing conceding to the majority decision, and, in situations where this occurred continually, people felt disempowered by the process.

In August 2019, the National Breast Imaging Academy programme inaugurated (*National Breast Imaging Academy* n.d.) an interprofessional breast education model, which may help to break down professional silos. Providing individuals with respect and constructive attitudes to work in a culture of interprofessional collaboration could alleviate traditional hierarchies (Green et al. 2017). While a vision of the NHS long term plan (NHS 2019) is a new leadership code that will cherish the required cultural values and behaviours, it is acknowledged that currently this is not established in some parts of the NHS. The results presented here suggest that in specific breast screening units there are unhealthy practice cultures (characterised by the absence of collegial support) and that hierarchical gradients, whether perceived or real, exist between the different professions.

The findings provide an insight into how and why guideline implementation might succeed in some settings and not in others. Future evaluation of all performance metrics with a uniform recording of third-person arbitrators (to include professional

role) and composition of group consensus may determine if there is a statistically significant difference dependent on these factors.

9.4 Issue 4. Planning the Service/Standardisation

A further issue identified from data triangulation in the current study was the concept of service planning and standardisation. Participants from all professional roles identified factors relating to the structure and design of services as a potential barrier to sustaining current performance outcomes. The most notable sub-theme was ‘silo' working, as multiple interviewees identified this, and it was additionally conveyed in free text survey comments. Fenwick, Seville, and Brunsdon (2009) concludes that silos have their origin in human behaviour. The healthcare landscape is radically changing, and as the breast radiology workforce is diminishing, it may be necessary to consider delivery of the national breast screening service using a different approach. The concept of centralisation/external arbitration was explored, to gain opinions on the benefits and disadvantages of an enhanced collaborative process. Instead of working in silos, several interviewees proposed a more inclusive approach which would work not only for arbitration cases but for the batch reading of screening mammograms to improve the use of existing resources. An approach which optimises services for a population rather than individual units may face some opposition, but this could offer a solution to units with a shortage of film readers. One of the most significant barriers when trying to break down silos is the mindset of healthcare professionals (Vatanpour, Khorramnia, and Forutan 2013). Interviewees expressed polarised views on external arbitration, and any suggested changes were often met with tentativeness and apprehension (Chapter 8.3.1.3).

Closer joint working at a regional and national level, with centralisation or collaboration of film reading/arbitration, may improve the quality of care, standardising practice and allowing units to benefit from each other's strengths. To create a more open mindset, it would be important that departments can see the advantages of a change in the infrastructure.

9.5 Issue 5 Digital infrastructure/ current technology

The final issue identified from integrating survey comments and qualitative interview data is a requirement for the existing technology to be updated and the prospects of new technology to deliver the transformational requirements of a modern health care system (NHS 2019).

The current NBSS reporting system is reported to be inflexible, creating inefficiency due to the reliance on paperwork. Functional deficiencies mean that there is a limited recording of data with no standardised way of documenting onto the system the people present during consensus (paper-trail). The most recent updates released in September 2019 include details of the type of abnormality for which a reader has recalled (mass, asymmetric density, microcalcification) and this information, in conjunction with known arbitrators, could help inform the ‘blind-spots' to which individual readers are prone. The laborious task of manually analysing combinations of consensus pair outcomes was described by respondents, and a system that could generate this information automatically may again help to define which combinations are optimal.

The proposition of a completely paperless system was explored with interviewees in the current study, with resulting divergent views. Topol (2019) states that:

'resistance to change and scepticism about technology are well-recognised barriers to progress."

Changes to the NBSS system that could facilitate a fully electronic system would need to be intuitive and straightforward. Interviewees stated that an electronic proforma would need to enhance and streamline the clinical workflow rather than adding to the work burden. This is in keeping with Moacdieh and Sarter (2015) who report that health IT tools which are not designed and employed to support the diagnostic process can detract from the clinician's reasoning activities, creating a cognitive burden and susceptibility to error. A potential area for future research is to investigate how AI algorithms can aid the identification of arbitration cancers by analysing large quantities of data to discover associations and trends in both reader performance and the type of mammographic abnormality that may not be apparent otherwise.

9.6 The Potential Role of New Technology

The final objective of this study was to comprehend the future role of new technology in breast screening reporting and arbitration practices. The findings presented here are a result of the integration of the literature and qualitative data.

9.6.1 Artificial Intelligence

A potential solution that may offset the human weaknesses in decision-making is to utilise the ability of computers and deep learning. Artificial Intelligence is an

emergent field, and therefore, there is a lack of literature regarding the clinical performance of these systems in a prospective clinical setting. There is a substantial potential risk associated with an AI algorithm as proven with the IBM Watson Health oncology algorithm, where numerous recommendations for treatment were incorrect (Quach 2018).

However, AI is increasingly recognised as having the promise to support and improve diagnostic performance, with the possibility of reducing diagnostic errors (El-Kareh, Hasan, and Schiff 2013). As the complexity of health care increases, radiology staff are subject to increasing amounts of data and volumes of images which some have argued may soon exceed human cognitive capacity (El-Kareh, Hasan, and Schiff 2013). With staffing shortages and increasing workloads, (The Royal College of Radiologists 2020) this is even more challenging. Error rates vary between individuals and may be due to a flawed human memory, variable disease presentation, and the heuristics and biases discussed in Chapter 3. Sokolovskaya et al. (2015) also state that as reporters are pressurised to work faster, there is an associated significant increase in the average interpretation error rate. In everyday clinical practice, high workloads may also be associated with distractions and interruptions, which are not conducive to effective perception (Waite et al. 2017, Donald and Barnard 2012).

AI is a disruptive technology that may improve productivity and confer economic benefits to deliver an NHS fit for the future (Topol, 2019). Although AI was not a specific focus within the interviews of the NHSBSP reporting staff in the current study, two interviewees raised this subject with opposing views. One deeming there

would be cases to challenge or defeat any AI system and the other proposing the use of an AI system to arbitrate.

Breast cancer screening represents an ideal application for AI as there are large datasets available for algorithm training and testing, and information (known outcomes) for validating clinical endpoints. The literature review and qualitative data (Clinical Director of a Med Tech company, Chapter 8.3.2.2) highlighted the substantial improvements over time in AI with deep learning algorithms. Early results using AI in other clinical settings such as retinal assessment (Gulshan et al. 2016), skin lesion analysis (Esteva et al. 2017, Phillips et al. 2019) and CT head scans (Chilamkurthy et al. 2018) demonstrate systems performing at a human specialist level capability. Kooi et al. (2017) report that breast cancer detection performance of AI systems are currently comparable to an average breast Radiologist, which may result in an improvement in the performance of breast cancer screening programmes (Trister, Buist, and Lee 2017). Conversely, Rodríguez-Ruiz et al. (2019) found that AI performed consistently lower than the best Radiologists in all datasets in their study. This is supported by a recent diagnostic accuracy study of screening mammograms (Schaffter et al. 2020) which found that while no solitary AI algorithm surpassed the Radiologists, a collective of AI algorithms with a single reader (Radiologist) demonstrated an increase in overall accuracy. Importantly, with double-reading and consensus as employed in many European countries, the authors report that the "addition of AI may not have as great an effect on improving overall diagnostic accuracy", but future research is required to train the AI

algorithms with consensus assessments which could improve consensus decisions alone.

More recently, McKinney and colleagues (2020) report the results from a deep learning AI system which correctly identified cancers with a similar degree of accuracy to expert Radiologists in screening mammograms. The authors also report a reduction of 5.7% and 1.2% (USA and UK, respectively) in false positives, and a decrease of 9.4% and 2.7% (USA and UK, respectively) in false negatives. The AI system was evaluated using a large-scale database of digital images (26,000 women/80,000 images) extracted from the UK NBSS, and a large enriched dataset from the USA (3,000 women). The authors undertook a simulation study utilising the AI system in the double-reading process and state that the

"AI system maintained non-inferior performance and reduced the workload of the second reader by 88%."

However, this was a research study, not a clinical study. The results are from images obtained mainly from a single manufacturer, and although the US readers may have utilised tomosynthesis, the results for each technology are not individually reported. The only demographic detail incorporated in the study is the age of the population, and Pisano (2020) states that the performance of an AI algorithms can be extremely reliant upon the population utilised in the training sets.

To date, there has been a narrow body of work comparing AI performance directly with humans. However, a study just published (Salim et al. 2020) is an external

evaluation of three commercially available AI CAD algorithms as independent mammography readers and assessed the screening performance when combined with radiologists. However, the AI CAD algorithms have not yet been approved by the US FDA to be used as an independent reader, and the vendors are unknown.

The progression of AI algorithms justifies the optimism that these systems can potentially aid reporters in several ways. Utilising the power of AI has the potential to tackle some of the current challenges in breast screening, including improving the accuracy of detection, the potential to increase efficiency, and expedite the detection of early cancers.

9.6.2 Role of AI in Detection and Decision Support.

Qualitative data (interview with the Med Tech Director, Chapter 8.3.2.2) disclosed that their software has a sensitivity and specificity higher than any existing Computer-Aided Detection (CAD) and stronger than a single expert breast Radiologist. A recent study (Rodríguez-Ruiz et al. 2019) compared the performance of Radiologists (breast cancer detection) reading unaided versus reading with an AI system and found that support from an AI system improved performance (measured by the area under the ROC curve), with no detrimental effect on reading times per case. However, this study was performed on a data set with a high proportion of cancers and therefore, not directly comparable with reporting in standard screening practice.

The most recent external validation study (Salim et al. 2020) used a screening mammography database (739 mammograms with breast cancer and 8066 randomly

sampled mammograms negative for breast cancer) to assess the performance of three AI algorithms in various settings. The number of true negative cases was increased (112 914) to give a rate of 6.5 cancers per 1000 when comparing the AI model performance with or without first and second readers, simulating a double-reading set-up. The authors conclude that

"*Combining the first reader with the best algorithm identified more cancer cases than combining the first and second readers".*

One of the algorithms is reported to demonstrate a diagnostic performance at the level of, or exceeding that of radiologists, with no marked benefit in performance when combining the three algorithms compared to using the best alone. Cancer detection rates were estimated to increase by 8% when added to the first reader result, but there was also an increase in the number of cases considered abnormal by 77% (true positives and false positives). Therefore, the benefit of replacing one human reader would result in a much larger number of cases and workload requiring arbitration review by a third human reader or consensus group. Interestingly, this study demonstrated that two human readers showed greater concordance (abnormal readings and false positives) than a human reader with an algorithm. This potentially supports the findings in the current study of bias when non-blind reading (Chapter 6.9.4.1 and Chapter 8.3.1.2).

Although the BSIS system may help to demonstrate specific strengths of readers, aiding optimal reader pairings, the present study's findings demonstrate that units do not necessarily have the luxury of sufficient staff to facilitate this. The distinct

advantage of an AI system to change the recall decision threshold by altering the operating point on the ROC curve is that this could be tailored to an individual reader's performance, allowing a high recaller to be paired with a low recall version of the algorithm. Importantly, (Rodriguez-Ruiz et al. 2019)stated that improved performance was attributed to an increase in the middle part of the ROC curve, implying that AI can enhance evaluation of indeterminate cases. This is significant clinically as the majority of arbitration cases are more likely to be equivocal. Rodríguez-Ruiz et al. (2019) also reported that the improvement in performance was greater for less experienced readers, but in the study by Watanabe et al. (2019) the very senior Radiologist who had the most mediocre results (CDR) showed the most improvement. Again, this represents a significant clinical finding as, with AI, newly qualified staff will not be influenced by dominant or senior individuals and pressure to concur with cultural norms, as found in the qualitative data presented this thesis.

Qualitative interview data emphasised that one concern of NBSS becoming a fully blind and paperless system was the potential for a reporter to inadvertently enter a normal/normal result when a recall was intended. AI systems can be set up to make a recall suggestion known (exam score and area marked); therefore, this has the potential to act as a safety net in this scenario.

An aim of The Five-Year Forward View (NHS 2014) was to reduce the care and quality gap striving to standardise high-quality care. The variability in current reader performance was highlighted as an issue in Chapter 2 (2.6.4). The use of AI algorithms could potentially augment poorer performance, reducing the variation in

the quality of decision-making, providing a more standardised service to women (Bell 2017).

9.6.3 Role of AI in Optimising Reading Strategies.

Qualitative analysis of survey responses (Chapter 6.9.5.1) demonstrates the variation in stratifying of cases dependent upon the level of suspicion of the abnormality. In some units, cases classified as suspicious are second read and automatically recalled for further assessment. Given the high workload and lack of experienced readers, the continued cost-effectiveness of double reading may in future be questioned. In particular, its value in cases when the first reporter has graded an abnormality as definitely malignant.

AI systems can indicate the risk of a cancer being present, based on a suspicion scoring of 1 - 10, with 10 representing a high risk of malignancy. This quantitative indicator may be useful in triaging and prioritising the worklist, with readers choosing to read the potential cancer cases first, to ensure timely recall to assessment. With no reported increase in reading time or even a potential decrease (Rodríguez-Ruiz et al. 2019), the benefits of AI support may enable more efficient reading. If readers are reassured when there is a low-risk score, they may spend less time on these cases and give more time to the most suspicious examinations.

A recent study by Lång et al. (2019), based on data from a sub-cohort of the Malmö Breast Tomosynthesis Screening Trial, assessed the performance of an AI algorithm in screening to determine if normal exams can be excluded, and the types of cancers the AI system did not detect. Seven invasive cancers were missed by the AI system,

six of which were visible by the assessing Radiologists. However, (Lång et al. 2019) demonstrated that:

· 19% of screening mammograms with a risk score of one or two could be removed from human screen reading without missing cancer. Furthermore, this reduced 5% of the false-positive cases.

· 69% of the screening mammograms were allocated a risk score of three to nine, with 31% confirmed as cancer. It was proposed that these could be single-read mammograms

· 12% of the screening mammograms were given a risk score of 10, and 69% of these were cancer. It was proposed that these are the examinations requiring a double read.

The above reading stratification resulted in a 54% reduction in workload. Rodríguez-Ruiz et al. (2019) offer an alternative proposition of a 50/50 pre-selection split with the 50% least suspicious read with one reader and AI and double reading for the 50% most suspicious cases. As there is an immediate second report, rather than waiting for the images to be second read by a human reader (later that day or another day), this offers the potential to speed up results.

McKinney et al. (2020), also explored how their AI system performed as a triage tool, using high-confidence operating points to automatically discard low-risk cases, to reduce the workload for reporters. Although the authors emphasised the potential of AI to sustain screening services in the face of workforce shortages, prospective

clinical trials will be required to benchmark and monitor performance and to evaluate in full the cost-effectiveness and extent of benefit to women in terms of longer-term outcomes. In an era where AI systems are coming close to performing human tasks, defining the level of human regulation required over an algorithm (based on the risk level) will be necessary, but foremost technology introduction will have to surmount concerns from healthcare professionals and the public (Johnson 2016). Introducing AI, as a stand-alone reader has yet to be studied, and regulations around the medicolegal consequences, if an AI system failed, would need to be established. Future studies in an actual screening scenario will validate and assess the real effect of AI support and will determine if an interactive system results in equivalence or improvements in reader performance (diagnostic performance and efficiency) and financial outcomes (Topol 2019). If there is equivalence and the cost is lower, and reporting is faster, then AI will be the 'dominant' technology.

9.6.4 AI as an Arbitrator

Gubern-Mérida (2019) has suggested that double reporting by humans continue and that AI be used as the third reader arbitrator. This represents an attractive proposal because AI algorithms could also offer a score on the likelihood of a lesion being cancer. The consensus group dynamic issues discussed in Chapter 4 (4.5) would be avoided, as would the inherent biases of an individual human reader. Potentially, this could help to reduce the false-positive rates and improve capacity on assessment clinics, although it would raise costs.

Overall, the prospects of AI in breast screening are considerable. Valuable consultant reading time could then be transferred to assessment clinics, follow-up

ultrasound scans following MRI findings, and Vacuum Excision Biopsies. At present, it is unknown if reporting and arbitration with an AI system would translate to fewer missed cancers and earlier diagnosis. Salim et al. (2020) report that in the analysis of interval cancers (diagnosed within 12 months of a negative screening mammogram) the best AI algorithm in their study achieved an area under the receiver operating characteristic curve (AUC) of 0.810. The authors, therefore, conclude that many of the prior mammograms will have malignant features present and that AI algorithms may aid earlier detection.

Conversely, the cost of the AI software and integration into the Picture Archiving and Communication System (PACS) workflow has to be considered. Potentially, if AI systems are used to augment a reader, there may be a potential increase in the number of assessment cases requiring consultant work-up. However, if the increased cases are true positives, this may save costs down the line by preventing more expensive treatment and increase the number of years of life, and, therefore, be cost-effective.

9.6.5 Further Potential Benefits of AI

Chapter 2 (2.8.1) emphasised that the use of DBT can increase the CDR of screening programmes, but this modality requires many images to be read; increasing the reading time by a factor of two (Gubern-Mérida 2019). If screening with DBT were to be implemented in the future, it would be imperative that the intelligent algorithms can constructively aid the reporter, by acting as a second reader (Harvey et al. 2019: 187). Marinovich et al. (2018) undertook a meta-analysis of thirteen

American studies and reported that the overall recall rate was 2.2% lower with DBT compared to DM examinations, with an associated increase in CDR of 1.1 per 1000 examinations. However, a significant finding from a recent US study (Sprague et al. 2020) is the considerable variability in results (CDR and recall rates) amongst Radiologists indicating that improvements in these performance metrics with DBT is not unanimous. The authors conclude that further research is required to assess the variability in Radiologist DBT screening performance. Future prospective studies may inform if 2D-mammography and AI have any clinical impact on the use of DBT, or whether AI trained on DBT images can increase the diagnostic performance further (Salim et al. 2020).

Also, discussed in Chapter 2 (2.3.4) was the impact of breast density as a risk factor, and the fact that the current BI-RADS breast density classification undertaken by reporters is subjective and hence variable. An algorithm capable of providing reproducible breast density measurements would be beneficial when exploring breast density to predict cancer risk. A recent study by Yala et al. (2019) reports the results of a mammography-based deep learning (DL) breast cancer risk model; the authors found that a hybrid model of mammograms in combination with traditional risk factors produced substantially improved risk prediction compared with the Tyrer-Cuzick model alone (version 8 which includes breast density). The ability to incorporate risk factors into the algorithms, for example, family history, and breast density was also confirmed by the Clinical Director of a Med Tech company, particularly when amalgamated with genomic data.

"So that is the next thing we are moving onto is breast density, and we already have a risk stratification algorithm much like Volpara, but we actually use that to actually inform our final call back decision" (Radiologist 4 Clinical Director of a Med Tech company).

AI models may, therefore, lead the way in potentially differentiating aggressive from indolent screen-detected cancers and thereby lessening the risk of over-diagnosis (Tice et al. 2005). Bahl et al. (2018) report a machine learning model that utilises established risk factors, text from the histology report, and results from a needle core biopsy to stratify patients diagnosed with a High-Risk Lesion (HRL). This offers the potential to differentiate HRL's that need a surgical excision and those (low risk of upgrade) that may require surveillance only, reducing unnecessary treatments. This possibility offers an opportunity for greater informed decision- making between the patient and the clinician and may assist in more personalised and precision patient care.

Gillies, Kinahan, and Hricak (2016) state that the future of radiology is the potential to transform a speciality of

"*qualitative interpretation to one of quantitative analysis".*

The incorporation of radiomics (quantitative measures of image texture) (Parekh and Jacobs 2017) to improve clinical decision-making could further pave the way to personalised precision health (Castaneda et al. 2015). However, research into this will require developing models that link a combination of imaging phenotypes from

large mammography datasets and genomic-level variables (Radio-genomics) (Houssami et al. 2017). Radio-genomics may uncover specific imaging biomarkers that could potentially help detect tumours that are often missed at mammography, such as diffuse invasive lobular cancers which often only exhibit subtle architectural distortions.

9.6.6 Current AI Activities Relevant to Breast Screening

In 2018, the East Midlands Radiology Consortium EMRAD created a partnership with two UK-based AI companies, to improve the development, testing and deployment of AI tools in the NHSBSP. The Test Bed project incorporates both clinical and operational processes aiming to

- optimise clinical service capacity
- improve patient care
- increase confidence in the use of machine learning tools.

The way AI is envisaged to work is as a first reader, then when the human reads, the AI opinion is available (the equivalent to non-blinded second reading).

An interview with the EMRAD Project Manager Simon Harris confirmed that the AI tool had been calibrated at two NHS sites, and the aim is to test the generalisability of the new deep learning mammography software. Simon is hoping to have completed the validation of the AI tool by the end of September 2020. The plan to move into prospective (real-world testing) is being drawn up. The tool will be implemented into the clinical workflow alongside the current double reading (for those women who have consented to have their mammograms read by AI). There

will be no change in their treatment, just an additional AI read in the background. At this stage, the AI outcomes will not be shared with the reporters. However, the data will be reviewed to ascertain if there is agreement or disagreement with the readers and to prove the accuracy of the tool. If the case goes to arbitration, the pilot will assess the AI result. The project team hope to conclude that their model is suitable for consideration as an independent reader in double-read screening programmes. Early results are reported to be 'really exciting' (Harris 2019). However, the 'holy grail' of a validated deep learning system that accurately makes recall decisions on par with, or superior to, human double-reading, while delivering explainable and interpretable results if required, may be some time away (Lehman 2020).

The critical dependency highlighted by Simon Harris is not to interrupt the clinical workflow. There are ongoing discussions with Hitachi who develop the NBSS software to ascertain how they can integrate this tool, together with PACs vendors regarding the imaging aspect. For AI to scale and spread throughout the breast screening units requires formal PHE agreement as safe practice. Simon Harris stated that 'getting through the regulations is the tricky part'.

9.6.7 AI as an Administrative Optimisation Tool

As discussed in Chapter 8 (8.3.1.3), breast screening sites currently operate in silos. AI-powered intelligent administrative systems could confer benefits to the breast screening programme through integration and improved planning, for instance, scheduling of clinics and staff resourcing. Capacity and demand planners may help to provide detailed forecasts and identify ways to alleviate these pressures. A

forewarning of increases in demand for cases to be recalled for assessment may help in proactively scheduling extra clinics before the peak in demand. The ability to estimate the impact of unplanned equipment downtime or workforce changes may help to reduce round length slippage. The facility to identify the optimal and less efficiently used clinics may also aid the planning of limited staff resources (Joshi and Morley 2019).

AI tools may also assist in the possibility of improving on patient DNA predictions which could be utilised to control 'smart clinic' features in NBSS. Joshi and Morley (2019) propose that similar methods may help to identify appointment slots which a client is most likely to accept, with the potential to increase first-time attendance rates, and thereby reduce the administration associated with clients re-booking their appointments for a more suitable time.

As considered in Chapter 8 (8.3.2.1), if the breast screening programmes specification and IT infrastructure evolved to support reporting and arbitration from other sites, further opportunities exist for client appointments. The utilisation of AI tools in this scenario could theoretically allow women to be screened at a location that incurs minimal travel time, such as closer to work, and the available service capacity. AI tools may also confer global benefits (Lehman 2020). The interview with the Clinical Director of a Med Tech company raised that opportunities could exist for countries that do not have breast screening resources.

"Third world countries for instance for us to just leapfrog them straight into having AI basic provisions and they're quite happy to do that because you know anything is better than nothing you know really in those countries".

This view is supported by a recent journal commentary, (Lehman 2020) stating that if AI models can differentiate between mammograms with and without cancer present,

"*Screening can be made available and affordable to a large population of women who currently have no access to the life-saving potential of quality screening".*

9.7 Evaluation of the Research Approach

The current research employed a mixed-methods approach to accomplish the aims, with the integration of the quantitative and qualitative data at multiple levels. Applying a mixed-methods approach in a unique setting was challenging to an early career researcher and required engaged and continuous analysis to identify and track outcomes from the literature (Lau and Kuziemsky 2017) with this research's empirical evidence. The limitations of the methods were explored in the respective chapters (5, 6 and 8). A strength of the study is that integration through the design was achieved via an explanatory sequential study. Methodological integration occurred through the process of connecting, in which the sampling frame for the interviews was selected from survey respondents ranked into high medium and low-performance parameters, and professional roles. Integration was also achieved at the interpretation and reporting stage with the quantitative and qualitative findings synthesised through this narrative using a weaving approach.

Although considerable research has been devoted to breast screening performance-based studies, these have used a purely quantitative approach (Taylor-Phillips et al. 2018, Blanks, Wallis, and Moss 1998, Bennett and Blanks 2007, Blanks et al. 2002, Burnside et al. 2018). Rather less attention has been paid to qualitative work, and this is the first study, to the researcher's knowledge, that has integrated quantitative and qualitative data on reporting/arbitration practices relating to breast screening. Using a mixed-methods approach in this setting was, therefore, highly novel and allowed an exploration of aspects which may have been missed using a purely quantitative approach.

9.7.1 Limitations of Methodology and Study Overall

A response from each breast screening unit was sought, but not every unit was represented. The findings are considered valuable regardless of being partial (for England), providing a foundation for accrual of further knowledge (Pawson 2013). Although the study had a broad geographical reach, the representativeness was most limited from the London region (33%). However, this did not limit the interview selection, as there were responses from each cell in the pre-determined sampling frame. It is acknowledged that the interviewees were respondents of the surveys, and although staff from all professional groups were included, it cannot be presumed that findings would be generalisable to all of the groups represented in the study. However, the results are still considered valid as in qualitative research; generalisability is not the ultimate aim (Scragg et al. 2017). This study revealed a lack of clarity in what participants classed to be arbitration or consensus. Hence although describing a group process, some respondents did not complete the team dynamics

element of the survey, resulting in a smaller sample of data (Burmeister and Aitken 2012). Survey respondents and non-respondents might have varied in ways that influence their team dynamics. However, this small sample demonstrated an understanding of the dynamics in breast consensus groups and revealed differences in perceptions amongst the varying professional roles.

The study provided an overview of barriers to implementation of the PHE guidance, but it is acknowledged that these may change over time. Nevertheless, this provides a valuable foundation from which to compare changes in future practice.

Further limitations of this study are identified, which could affect the results. Regarding the performance data; the individual round length for each unit is unknown, and hence any slippage could affect a units SDR. The KC62 data (publicly available) utilised for this study does not provide data on non-invasive or micro-invasive disease for individual units. It was therefore not possible to ascertain if higher recall rates were associated with higher DCIS rates. The reading practices, arbitration strategies and Radiographer arbitration data could only be analysed for the 49/80 (61%) of respondent units. An analysis of data for all 80 units may show different and statistically significant results. Some units introduced Radiographer arbitration within 2016-2017 and therefore a review of national data over an extended period would need to be evaluated to ascertain if there is any impact on overall recall rates.

The qualitative findings presented in this thesis are a result of data analysis and interpretation by an individual researcher. It is recognised that there are issues of

validity and reliability in qualitative research (Noble and Smith 2015). The study design mitigated against these where possible by checking face validity and, where possible, interpretation with a second independent researcher.

Section 2: Implications of the study findings

The findings of this research have led to the following key organisational and national recommendations.

9.8 Recommendations

Recommendation 1: Blinded Double Reading

Blind independent double reading should be considered to obtain the best insight into individual reader performance and standardisation of practice.

Analysis of qualitative data (Chapter 6.9.4.1 survey comments and Chapter 8.3.1.2) reveals the concept of 'following-on' for some newly qualified readers with a modification of their reporting practices and a subsequent homogenisation of reading. True blind second reading would allow improved monitoring of readers as this will provide more substantial amounts of unbiased data. There is limited literature in this field, but with less invasive interval cancers and a slightly higher CDR detected with blind double reading compared to non-blinded reading (Klompenhouwer et al. 2015a) a change in practice may have the potential to improve the sensitivity of a programme. The caveat is that true blind reading may increase the number of cases requiring arbitration.

Recommendation 2: Blinded Arbitration/Consensus

To obtain independent non-biased opinions on arbitration cases consideration should be given to anonymisation of which individuals have reported/recalled the case.

Analysis of qualitative data (survey comments Chapter 6.9.4.1 and Chapter 8.3.1.2) reveals in some units a biased perspective in arbitration. Rather than making an independent opinion on a potential abnormality and arbitrating effectively, some individuals conveyed their judgement was based on knowing the professional who recalled the case. Anonymised arbitration or consensus (individual identities are not known) would reduce several biases and may have the potential to reduce recall rates in some units where dominant individuals who are a high recaller control practice. Conversely, it may prevent erroneous dismissal of a positive case when the credibility of high recaller is overridden. The literature review highlighted that prestige and dominant hierarchies develop over time within groups and may be difficult to change. Human decision-making is susceptible to psychological biases, the impact of which may be considerable. Triangulation of qualitative data and the literature indicates that the application of Artificial Intelligence has the potential to support and improve clinical decision-making both within the reporting and arbitration setting (Lång et al. 2019).

Recommendation 3: Careful Selection of Arbitrators

Directors of breast screening should consider using the BSIS data to support delegation of solitary third reader arbitration/lead of consensus review meetings.

Analysis of qualitative data (Chapter 6 survey comments 6.9.4.2 and Chapter 8, sub-theme 3.1) emphasises that expertise in screen reading is characterised by the sensitivity and specificity of the reporter. Although first/second reader BSIS data may not transfer directly to arbitration, selecting individuals with the highest sensitivity and specificity may be more effective compared to the current situation, where the task is predominantly based on professional role. Substantiation of third reader outcome measures by professional role would be a valuable extension of the study.

Qualitative findings (Survey comments Chapter 6.9.10 and Chapter 8 sub-theme 2.1, 3.2, and 3.3) also reinforce that feedback on performance (true positive and false negative cases) is integral in developing expertise, inducing reflective learning, corroborating accurate reasoning and developing confidence. This is in keeping with the substantive theory proposed by Ericsson (2004) regarding expert 'deliberate practice' as a method which could improve the reasoning processes of Radiographers as they progress from a novice to an expert practitioner.

Developing clinical reasoning skills are essential for reporters who have completed the trainee phase, and continuing education can be leveraged to progress these skills as a core aim (Cruz, Pimenta, and Lunney 2009). Until recently, the non-

technical skills of performance, for instance, clinical leadership were not focused on in the curricula and training for radiographers. As a result, there will be experienced advanced and consultant practitioners who have not been equipped with the expertise or feel empowered to undertake leadership roles. Preceptorship periods may be useful for individuals lacking confidence in undertaking the role. The caveat is that delegation to specific individuals may create unease between peers and different professional roles.

Qualitative survey comments (Chapter 6.9.10) also indicate that there are limited opportunities for feedback on arbitration performance in some units. NHSBSP reporters previously attended regional interval cancer reviews, but these are no longer provided. A potential area for future research would be to develop an electronic (web-based) programme to review cancers that were arbitrated to routine recall. Analysis of the imaging characteristics (mammographic abnormality and visualisation), review of prior mammography, and discrepant reads would be a valuable educational tool. A web-based programme would allow accessibility and participation from a wider staff cohort, providing a forum for continuous learning. Reporters will review their local interval cancers (including arbitrated interval cancers) and false-negative cases and may be fixated on the missed cancers. However, it is also essential to review cases that are recalled and subsequently returned to routine recall at arbitration or assessment. A review of these cases may have a more significant impact on reducing the recall rate. In a learning healthcare system, it is recommended that organisations adopt policies and practices to deliver

systematic feedback and open discussion in a non-punitive culture for service improvement.

Recommendation 4: Software Design/Update

Consideration should be given to secure funding for rewrites within the National Breast Screening Computer System (NBSS) to improve usability, enable optimisation of consensus groups and to facilitate true blind reading/arbitration.

Analysis of qualitative data (Chapter 6.9.4.1 survey comments and Chapter 8.3.2.1 interview data) highlights that the current reporting system is deemed by all professional roles to be suboptimal, lacking sophistication and reliant upon paperwork. Specific clinical trials (PROSPECTS) require blind reading but a transition to a paperless system would be necessary to support this wholeheartedly (requiring anonymisation of results and readers with an electronic recall proforma). Design recommendations would need to be explored to support this as the system must not detract from clinical efficiency.

The findings from this study suggest that in the main, an electronic recall proforma would be well received, but there is caution regarding losing a paper system that currently acts as a safety net. Therefore, further development and research is required to explore this potential. Hitachi the health IT vendor for NBSS would need to collaborate with users to identify optimal practices in the design and implementation to ensure usability and streamline the clinical workflow.

Recommendation 5: Centralisation/Independent Arbitration

Consideration should be given to independent (but internal to the NHSBSP) arbitration or consensus.

Triangulation of qualitative, quantitative data and the literature has demonstrated the difficulties in defining performance metrics for arbitration as currently, the numbers generated in a year are too small to judge. However, qualitative data (Chapter 8, sub-theme 1.3) reveals that centralisation could be a powerful tool for normalising arbitration across the country. It would make it possible, for the first time, for enough arbitrations to be performed by an individual to allow performance to be monitored more accurately. Analysis of qualitative data and the literature review (Buist et al. 2014, Carney et al. 2013, Miglioretti et al. 2007, and Steel 2016) emphasised the current variability (regional variation quoted as 25%-75% of arbitrations recalled for the same client) in individual third reader performance and the issue in specific units of consensus with dominant individuals overriding. This could be problematic if the reader is not specific as this will impact on the recall rate. With further progression of AI, it may also be possible to look at variation research in datasets to identify inconsistencies across regions and explore undetected trends.

The caveat to independent arbitration is that a robust electronic infrastructure would be required to support remote arbitration. There would be a requirement for the IT systems to support efficient and effective transfer of client information across units to facilitate the diagnostic process (with no unanticipated downtime) to avoid introducing a delay into a programme that is tightly governed by time constraints.

As IT is progressively incorporated into health care, clinicians will have the potential to facilitate external diagnostic decision-making, communication and collaboration with peers (Thibault 2013). However, qualitative findings highlight that some staff might perceive they are losing the skill set and expertise of decision-making in challenging cases or find it difficult to accept an opinion from an external arbitrator. Nevertheless, if collaborative practice models can provide an efficient, reliable system that leads to improved patient outcomes they should be considered.

Section 3: Suggestions for Further research

While this research addressed the study aims, it has created several additional research questions.

9.9 Further Evaluation.

One output from this research was the analysis of performance metrics based on the professional role of the arbitrator. It would be valuable to establish the impact of third reader arbitration on a broader scale (all 80 units in the country); in particular, to measure quantitative outcomes. A uniform recording of third-person arbitrators (to include professional role) and composition of group consensus may determine if there is a statistically significant difference dependent on these factors. The following research questions need to be answered: Is there any effect on the performance metrics in units where Radiographers are undertaking third reader arbitration? Do Radiographers initiate more recalls to assessments than Radiologists in this setting?

The current thesis has suggested that arbitration is a vital part of the service that little attention is paid to, specifically in choosing who arbitrates and how units monitor the outcomes. Given the potential number of cancers in an arbitration pile, it is an important question. A trial study (with the true incidence of cancer) to evaluate individual performance may help to inform the selection of third reader arbitrators. Re-assessment every few years would be required, but this would provide sufficient numbers to allow continuous monitoring of how third reader arbitration works and potentially improved outcomes, and improved learning for units. However, it is acknowledged that there is contradictory evidence on whether performance in reading test sets correlates to clinical performance (Scott et al. 2009, Rutter and Taplin 2000, Soh et al. 2015). The current PERFORMS test does not include prior imaging. An arbitration test set with previous mammograms (if an incident screen) would better replicate the clinical screening setting, and the potential to use adaptive tests Computerized Adaptive Testing (CAT) using iterative algorithms may be useful as the level of difficulty of subsequent cases are based on correct or incorrect decisions on the prior case.

As proposed by Wolf et al. (2015) and Barnett et al. (2019) a collective intelligence study in which several independent decisions on a case are pooled could provide a constructive approach to consensus negating the adverse outcomes associated with group dominance. The accuracy in a clinical setting is unknown but could warrant further study to establish the impact on recall rates and CDR.

To establish the benefits of complete blind reading versus non-blind reading would first require updates to the NBSS reporting system. A randomised controlled trial

may then establish the impact on the number of arbitration cases and may establish more significant differences in reading ability.

No previous studies were identified that addressed how teamwork or the organisational environment impacts on the clinical decision-making strategies in breast screening. This study provided new information on the impact of organisational culture on team decision-making, and the data suggests that there is the potential for improvement in some units. As the results did not demonstrate an association between mean recall rates, four-year average SDR (prevalent and incident) and small cancer detection rates (prevalent and incident) for all variables, this may suggest that human factors explain the variation in unit performance. Human factor characteristics associated with performance include leadership, communication, decision making, teamwork and workload management (Catchpole 2013). Future research efforts could help to understand these complex relationships and power structures.

A research study mapping the 'perfect team' based on low recall rates and high CDR may help understand the complex inter-relationship between the staff (people and culture) and the organisational context (systems and processes). This may identify ways of working that can influence quality improvement in teams, support cultural change, and optimise the process. Human factor analysis within these teams may help to elucidate the limitations and unreliability of human performance. Exploring current practice's reliability and efficiency may improve the human clinical potential and the contribution that standardisation of practice could bring in this setting. A more specific focused tool may be valuable in evaluating team dynamics and

determining cultural and organisational characteristics that improve diagnosis, support effective teamwork and help to identify best practices.

This literature review also identified that there is an ongoing international consideration surrounding the cost-effectiveness of double reading since the transition to digital mammography. This is particularly pertinent if, as this study implies, non-blinded reading/arbitration is the norm, with the associated decision-making biases. Taylor-Phillips and colleagues (2018) state that a randomised controlled trial would be necessary to comprehend any differences in outcomes between single reading and double reading programmes. If AI is introduced to support single reading, future research may provide an understanding of the human-automation interaction assessing any improvements in reader performance, or reader errors which may have been influenced by erroneous scores from the AI system.

9.10 Overall Conclusions and Contributions to the Existing Literature

There is a paucity of research studies within breast screening arbitration practices. The complexities and multiple variations discussed have subsequently made it problematic to establish an evidence base for best practice.

The quality of breast cancer diagnosis may differ with factors such as the volume of mammograms reported, experience, workload and time pressures. As discussed in Chapter 8 strategies that may improve patient outcomes include (i) access to expertise and technologies (AI), (ii) improved education and peer learning, and (iii) innovative models of service delivery.

In summary, this thesis has consequently produced the following original contributions:

1. A detailed understanding of current reporting and arbitration processes in breast screening with identification and analysis of factors relative to specific performance metrics.

2. Practice recommendations regarding blind reading/arbitration to provide improved film reader data profiles and standardisation

3. Recommendations surrounding updates to the national breast screening reporting system (NBSS) to facilitate blind reading/arbitration and a transition to a paperless system

4. Exploration of the PHE arbitration guidance and its impact on respondent units; highlighting some differences between practicalities and aspirations.

5. Provided considerations surrounding alternative models of service delivery.

9.11 Concluding Remarks

This research has demonstrated the complexity of factors associated with reporting and arbitration practices in a breast screening setting. A greater understanding of how these various factors influence the recall rate is fundamental in optimising clinical practice.

A quote from a participant summarises the current situation on arbitration

"I think it's hidden, but it's in danger of it being a really key part of the service that we don't pay any attention to. There has never been that much work done… and it's something that to some extent we should really be turning our attention to".

Behaviour change is recognised as the most significant challenge when attempting to influence a change in a current healthcare system (Braithwaite 2018). Human decision-making will always have associated flaws (El-Kareh, Hasan, and Schiff 2013, Schiff et al. 2009). The use of future technology and innovative models of service delivery may help to standardise practice and improve outcomes.

The breast cancer screening programme was mainly suspended in the UK by the coronavirus pandemic. With the possibility of further waves of COVID-19 cases, it is essential for units to continually adapt to keep services running as much as possible whilst maintaining safety. Before the COVID-19 crisis, there were unfilled breast screening posts across the country, and this will become more pressing during the pandemic. Services will have to plan how they will catch up the backlog, and the capacity they will need. While staff are vital to the recovery, the crisis has also compelled services to find new and innovative ways of working, by modifying workflow and adopting technology, for example, Multidisciplinary meetings performed online. AI could potentially tackle some of the current challenges in breast screening, but further research is needed on optimising human/AI decision-making, via rigorous prospective clinical trials.

References

Abbott, P., Sapsford, R., and Sapsford, R. (1998) *Research Methods for Nurses and the Caring Professions*. 2nd edn. Open University Press

AbuAlRub, R.F., Gharaibeh, H.F., and Bashayreh, A.E.I. (2012) 'The Relationships Between Safety Climate, Teamwork, and Intent to Stay at Work Among Jordanian Hospital Nurses'. *Nursing Forum* 47 (1), 65–75

Adams, K., Greiner, A.C., and Corrigan, J.M. (2004) *1st Annual Crossing the Quality Chasm Summit*. Washington, D.C.: National Academies Press

AGREE Collaboration (2003) 'Development and Validation of an International Appraisal Instrument for Assessing the Quality of Clinical Practice Guidelines: The AGREE Project'. *Quality and Safety in Health Care* 12 (1), 18–23

Alberdi, E., Povykalo, A., Strigini, L., and Ayton, P. (2004) 'Effects of Incorrect Computer-Aided Detection (CAD) Output on Human Decision-Making in Mammography.' *Academic Radiology* 11 (8), 909–918

Alberdi, R.Z., Llanes, A.B.F., Ortega, R.A., Expósito, R.R., Collado, J.M.V., Verdes, T.Q., Ramos, C.N., Sanz, M.E., Trejo, D.S., Oliveres, X.C., and Group(1), and the C. (Cumulative F.P.R. (2011) 'Effect of Radiologist Experience on the Risk of False-Positive Results in Breast Cancer Screening Programs'. *European Radiology* 21 (10), 2083–2090

Allen, S.W., Brooks, L.R., Norman, G.R., and Rosenthal, D. (1988) 'Effect of Prior

Examples on Rule-Based Diagnostic Performance.' *Research in Medical Education : Proceedings of the ... Annual Conference. Conference on Research in Medical Education* 27, 9–14

ALQahtani, D.A., Rotgans, J.I., Mamede, S., ALAlwan, I., Magzoub, M.E.M., Altayeb, F.M., Mohamedani, M.A., and Schmidt, H.G. (2016) 'Does Time Pressure Have a Negative Effect on Diagnostic Accuracy?' *Academic Medicine* 91 (5), 710–716

Anderson, R.A., Crabtree, B.F., Steele, D.J., and McDaniel, R.R. (2005) 'Case Study Research: The View From Complexity Science'. *Qualitative Health Research* 15 (5), 669–685

Anderson, R.A., Plowman, D., Corazzini, K., Hsieh, P.-C., Su, H.F., Landerman, L.R., and McDaniel, R.R. (2013) 'Participation in Decision Making as a Property of Complex Adaptive Systems: Developing and Testing a Measure'. *Nursing Research and Practice* 2013, 706842

Anttinen, I., Pamilo, M., Soiva, M., and Roiha, M. (1993) 'Double Reading of Mammography Screening Films--One Radiologist or Two?' *Clinical Radiology* 48 (6), 414–21

Argyris, C. and Schön, D.A. (1978) *Organizational Learning*. Addison-Wesley Pub. Co

Arrow, K.J., Forsythe, R., Gorham, M., Hahn, R., Hanson, R., Ledyard, J.O., Levmore, S., Litan, R., Milgrom, P., Nelson, F.D., Neumann, G.R., Ottaviani, M., Schelling, T.C., Shiller, R.J., Smith, V.L., Snowberg, E., Sunstein, C.R., Tetlock, P.C., Tetlock,

P.E., Varian, H.R., Wolfers, J., and Zitzewitz, E. (2008) 'ECONOMICS: The Promise of Prediction Markets'. *Science* 320 (5878), 877–878

Atif, A., Richards, D., and Bilgin, A. (2012) 'Estimating Non-Response Bias in a Web-Based Survey of Technology Acceptance: A Case Study of Unit Guide Information Systems'. in *Australasian Conference on Information Systems*. held 2012. 1–10

Autier, P. and Boniol, M. (2018) 'Mammography Screening: A Major Issue in Medicine'. *European Journal of Cancer* 90, 34–62

Baars, B. (1980) *The Competing Plans Hypothesis: A Heuristic Viewpoint on the Causes of Speech ErrorsTemporal Variables in Speech : Studies in Honour of Frieda Goldman-Eisler.* De Gruyter Mouton, 13–19

Bago, B. and De Neys, W. (2017) 'Fast Logic?: Examining the Time Course Assumption of Dual Process Theory'. *Cognition* 158, 90–109

Bahl, M., Barzilay, R., Yedidia, A.B., Locascio, N.J., Yu, L., and Lehman, C.D. (2018) 'High-Risk Breast Lesions: A Machine Learning Model to Predict Pathologic Upgrade and Reduce Unnecessary Surgical Excision'. *Radiology* 286 (3), 810–818

Baird, M. (2008) 'Towards the Development of a Reflective Radiographer: Challenges and Constraints'. *Biomedical Imaging and Intervention Journal* 4 (1), e9

Baker, D.P., Day, R., and Salas, E. (2006) 'Teamwork as an Essential Component of High-Reliability Organizations.' *Health Services Research* 41 (4 Pt 2), 1576–98

Bakwin, H. (1945) 'Pseudodoxia Pediatrica'. *New England Journal of Medicine* 232 (24), 691–697

Balogh, E.P., Miller, B.T., Ball, J.R., Care, C. on D.E. in H., Services, B. on H.C., Medicine, I. of, The National Academies of Sciences, Engineering, and M., and National Academies of Sciences, Engineering, and M. (2015) *Improving Diagnosis in Health Care*. ed. by Balogh, E.P., Miller, B.T., and Ball, J.R. Washington, D.C.: National Academies Press

Bang, D. and Frith, C.D. (2017) 'Making Better Decisions in Groups.' *Royal Society Open Science* 4 (8), 170193

Bankier, A.A., Levine, D., Halpern, E.F., and Kressel, H.Y. (2010) 'Consensus Interpretation in Imaging Research: Is There a Better Way?' *Radiology* 257 (1), 14–17

Banning, M. (2008) 'A Review of Clinical Decision Making: Models and Current Research'. *Journal of Clinical Nursing* 17 (2), 187–195

Barnes, B.B.E., Steindorf, K., Hein, R., Flesch-Janys, D., and Chang-Claude, J. (2011) 'Population Attributable Risk of Invasive Postmenopausal Breast Cancer and Breast Cancer Subtypes for Modifiable and Non-Modifiable Risk Factors'. *Cancer Epidemiology* 35 (4), 345–352

Barnett, M.L., Boddupalli, D., Nundy, S., and Bates, D.W. (2019) 'Comparative Accuracy of Diagnosis by Collective Intelligence of Multiple Physicians vs Individual Physicians'. *JAMA Network Open* 2 (3), e190096 1-11

Barrows, H.S., Robyn, M., Tamblyn, N., and Tamblyn, R.M. (1980) *Problem-Based Learning An Approach to Medical Education*. 1st edn. Springer Pub. Co

Bell, D.E., Raiffa, H., and Tversky, A. (1988) 'DESCRIPTIVE, NORMATIVE, AND PRESCRIPTIVE INTERACTIONS IN DECISION MAKING'. in *Decision Making*. ed. by Bell, D.E., Raiffa, H., and Tversky, A. Cambridge: Cambridge University Press, 9–30

Bell, J. (2017) *Life Sciences Industrial Strategy-A Report to the Government from the Life Sciences Sector*.

Bellio, G., Marion, R., Giudici, F., Kus, S., Tonutti, M., Zanconati, F., and Bortul, M. (2017) 'Interval Breast Cancer Versus Screen-Detected Cancer: Comparison of Clinicopathologic Characteristics in a Single-Center Analysis'. *Clinical Breast Cancer* 17 (7), 564–571

Bennett, R.L. and Blanks, R.G. (2007) 'Should a Standard Be Defined for the Positive Predictive Value (PPV) of Recall in the UK NHS Breast Screening Programme?' *The Breast* 16 (1), 55–59

Bennett, R.L., Sellars, S.J., Blanks, R.G., and Moss, S.M. (2012) 'An Observational Study to Evaluate the Performance of Units Using Two Radiographers to Read

Screening Mammograms'. *Clinical Radiology* 67 (2), 114–121

Berbaum, K.S., Schartz, K.M., Caldwell, R.T., Madsen, M.T., Thompson, B.H., Mullan, B.F., Ellingson, A.N., and Franken, E.A. (2013) 'Satisfaction of Search from Detection of Pulmonary Nodules in Computed Tomography of the Chest'. *Academic Radiology* 20 (2), 194–201

Berenson, R.A., Upadhyay, D.K., and Kaye, D.R. (2014) *Placing Diagnosis Errors on the Policy Agenda Timely Analysis of Immediate Health Policy Issues April 2014*.

Berger, R. (2015) 'Now I See It, Now I Don't: Researcher's Position and Reflexivity in Qualitative Research'. *Qualitative Research* 15 (2), 219–234

Berlin, L. (2007) 'Radiologic Errors and Malpractice: A Blurry Distinction'. *American Journal of Roentgenology* 189 (3), 517–522

Bernardi, D., Macaskill, P., Pellegrini, M., Valentini, M., Fantò, C., Ostillio, L., Tuttobene, P., Luparia, A., and Houssami, N. (2016) 'Breast Cancer Screening with Tomosynthesis (3D Mammography) with Acquired or Synthetic 2D Mammography Compared with 2D Mammography Alone (STORM-2): A Population-Based Prospective Study.' *The Lancet. Oncology* 17 (8), 1105–1113

Bhutta, Z., Lassi, S., and Mansoor, N. (2011) *Systematic Review on Human Resources for Health Interventions to Improve Maternal Health Outcomes: Evidence from Developing Countries.* Karachi Pakistan Division of Women and Child Health Aga Khan University 2011.

Bird, C.M. (2005) ‘How I Stopped Dreading and Learned to Love Transcription’. *Qualitative Inquiry* 11 (2), 226–248

Blanca, M.J., Alarcón, R., Arnau, J., and Bendayan, R. (2017) ‘Non-Normal Data: Is ANOVA Still a Valid Option?’ *Psicothema* 29 (4), 552–557

Blandford, A.E. (2013) ‘Semi-Structured Qualitative Studies’. in *The Encyclopedia of Human-Computer Interaction*. 2nd edn.

Blanks, R.G. (2011) ‘Commentary on: Breast Screening Policy: Are We Heading in the Right Direction?’ *Clinical Radiology* 66 (10), 920–921

Blanks, R.G., Bennett, R.L., Wallis, M.G., and Moss, S.M. (2002) ‘Does Individual Programme Size Affect Screening Performance? Results from the United Kingdom NHS Breast Screening Programme’. *Journal of Medical Screening* 9 (1), 11–14

Blanks, R.G., Given-Wilson, R.M., Cohen, S.L., Patnick, J., Alison, R.J., and Wallis, M.G. (2019) ‘An Analysis of 11.3 Million Screening Tests Examining the Association between Recall and Cancer Detection Rates in the English NHS Breast Cancer Screening Programme’. *European Radiology*

Blanks, R.G., Wallis, M.G., and Moss, S.M. (1998) ‘A Comparison of Cancer Detection Rates Achieved by Breast Cancer Screening Programmes by Number of Readers, for One and Two View Mammography: Results from the UK National Health Service Breast Screening Programme’. *Journal of Medical Screening* 5 (4), 195–

Blumenthal-Barby, J.S. and Krieger, H. (2015) 'Cognitive Biases and Heuristics in Medical Decision Making'. *Medical Decision Making* 35 (4), 539–557

van Bodegom-Vos, L., Verhoef, J., Dickmann, M., Kleijn, M., van Vliet, I., Hurkmans, E., van der Wees, P., and Vliet Vlieland, T. (2012) 'A Qualitative Study of Barriers to the Implementation of a Rheumatoid Arthritis Guideline Among Generalist and Specialist Physical Therapists'. *Physical Therapy* 92 (10), 1292–1305

Boldt, A., Schiffer, A.M., Waszak, F., and Yeung, N. (2019) 'Confidence Predictions Affect Performance Confidence and Neural Preparation in Perceptual Decision Making'. *Scientific Reports* 9 (1), 1–17

Bonikowski, B. (2016) 'Nationalism in Settled Times'. *Annual Review of Sociology* 42 (1), 427–449

Bornstein, B.H. and Christine Emler, A. (2001) 'Rationality in Medical Decision Making: A Review of the Literature on Doctors' Decision-Making Biases'. *Journal of Evaluation in Clinical Practice* 7 (2), 97–107

Bornstein, B.H., Emler, C., and Chapman, G.B. (1999) 'Rationality in Medical Treatment Decisions: Is There a Sunk-Cost Effect?' *Social Sciences and Medicine* 49 (2), 215–222

Bovens, L and Hartmann, S. (2003) *Bayesian Epistemology*. Oxford: Oxford University

Press

Boyd, N.F., Byng, J.W., Jong, R.A., Fishell, E.K., Little, L.E., Miller, A.B., Lockwood, G.A., Tritchler, D.L., and Yaffe, M.J. (1995) 'Quantitative Classification of Mammographic Densities and Breast Cancer Risk: Results From the Canadian National Breast Screening Study'. *JNCI Journal of the National Cancer Institute* 87 (9), 670–675

Boyd, N.F., Guo, H., Martin, L.J., Sun, L., Stone, J., Fishell, E., Jong, R.A., Hislop, G., Chiarelli, A., Minkin, S., and Yaffe, M.J. (2007) 'Mammographic Density and the Risk and Detection of Breast Cancer'. *New England Journal of Medicine* 356 (3), 227–236

Boynton, P.M. and Greenhalgh, T. (2004) 'Selecting, Designing, and Developing Your Questionnaire.' *BMJ (Clinical Research Ed.)* 328 (7451), 1312–5

Brace, I. (2005) 'Questionnaire Design : How to Plan, Structure and Write Survey Material for Effective Market Research'. *Choice Reviews Online* 42 (6), 3520

Brady, A.P. (2017) 'Error and Discrepancy in Radiology: Inevitable or Avoidable?' *Insights into Imaging* 8 (1), 171–182

Braithwaite, J. (2018) 'Changing How We Think about Healthcare Improvement.' *BMJ (Clinical Research Ed.)* 361, k2014

Brand, J.S., Czene, K., Shepherd, J.A., Leifland, K., Heddson, B., Sundbom, A., Eriksson,

M., Li, J., Humphreys, K., and Hall, P. (2014) 'Automated Measurement of Volumetric Mammographic Density: A Tool for Widespread Breast Cancer Risk Assessment.' *Cancer Epidemiology, Biomarkers & Prevention : A Publication of the American Association for Cancer Research, Cosponsored by the American Society of Preventive Oncology* 23 (9), 1764–72

Brannick, T. and Coghlan, D. (2007) 'In Defense of Being "Native": The Case for Insider Academic Research'. *Organizational Research Methods* 10 (1), 59–74

Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R.L., Torre, L.A., and Jemal, A. (2018) 'Global Cancer Statistics 2018: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries'. *CA: A Cancer Journal for Clinicians* 68 (6), 394–424

Brekhus, W. (2015) *Culture and Cognition : Patterns in the Social Construction of Reality*. 1st edn. Polity Press

Brennan, S.E., Bosch, M., Buchan, H., and Green, S.E. (2013) 'Measuring Team Factors Thought to Influence the Success of Quality Improvement in Primary Care: A Systematic Review of Instruments'. *Implementation Science* 8 (1), 20

Brodersen, J. and Siersma, V.D. (2013) 'Long-Term Psychosocial Consequences of False-Positive Screening Mammography.' *Annals of Family Medicine* 11 (2), 106–15

Brooks, L.R., Norman, G.R., and Allen, S.W. (1991) 'Role of Specific Similarity in a

Medical Diagnostic Task.' *Journal of Experimental Psychology. General* 120 (3), 278–87

Broustas, C.G. and Lieberman, H.B. (2014) 'DNA Damage Response Genes and the Development of Cancer Metastasis.' *Radiation Research* 181 (2), 111–30

Brown, J., Bryan, S., and Warren, R. (1996) 'Mammography Screening: An Incremental Cost Effectiveness Analysis of Double versus Single Reading of Mammograms'. *BMJ* 312 (7034), 809–812

Brush, J.E., Sherbino, J., and Norman, G.R. (2017) 'How Expert Clinicians Intuitively Recognize a Medical Diagnosis'. *The American Journal of Medicine* 130 (6), 629–634

Bryman, A. (2015) *Social Research Methods*. 5th edn. Oxford University Press

Bryman, A., Becker, S., and Sempik, J. (2008) 'Quality Criteria for Quantitative, Qualitative and Mixed Methods Research: A View from Social Policy'. *International Journal of Social Research Methodology* 11 (4), 261–276

Buist, D.S.M., Anderson, M.L., Haneuse, S.J.P.A., Sickles, E.A., Smith, R.A., Carney, P.A., Taplin, S.H., Rosenberg, R.D., Geller, B.M., Onega, T.L., Monsees, B.S., Bassett, L.W., Yankaskas, B.C., Elmore, J.G., Kerlikowske, K., and Miglioretti, D.L. (2011) 'Influence of Annual Interpretive Volume on Screening Mammography Performance in the United States'. *Radiology* 259 (1), 72–84

Buist, D.S.M., Anderson, M.L., Smith, R.A., Carney, P.A., Miglioretti, D.L., Monsees, B.S., Sickles, E.A., Taplin, S.H., Geller, B.M., Yankaskas, B.C., and Onega, T.L. (2014) 'Effect of Radiologists' Diagnostic Work-up Volume on Interpretive Performance'. *Radiology* 273 (2), 351–364

Burkell, J. (2003) *The Dilemma of Survey Nonresponse*. vol. 25

Burmeister, E. and Aitken, L.M. (2012) 'Sample Size: How Many Is Enough?' *Australian Critical Care* 25 (4), 271–274

Burns, K.E.A., Duffett, M., Kho, M.E., Meade, M.O., Adhikari, N.K.J., Sinuff, T., Cook, D.J., and ACCADEMY Group (2008) 'A Guide for the Design and Conduct of Self-Administered Surveys of Clinicians'. *Canadian Medical Association Journal* 179 (3), 245–252

Burnside, E.S., Vulkan, D., Blanks, R.G., and Duffy, S.W. (2018) 'Association between Screening Mammography Recall Rate and Interval Cancers in the UK Breast Cancer Service Screening Program: A Cohort Study'. *Radiology* 288 (1), 47–54

Bushe, G.R. and Chu, A. (2011) 'Fluid Teams: Solutions to the Problems of Unstable Team Membership'. *Organizational Dynamics* 40, 181–188

Cabana, M.D., Medzihradsky, O.F., Rubin, H.R., and Freed, G.L. (2001) 'Applying Clinical Guidelines to Pediatric Practice.' *Pediatric Annals* 30 (5), 274–82

Cachia, M. and Millward, L. (2011) 'The Telephone Medium and Semi-structured

Interviews: A Complementary Fit'. *Qualitative Research in Organizations and Management: An International Journal* 6 (3), 265–277

Campbell, W.K., Goodie, A.S., and Foster, J.D. (2004) 'Narcissism, Confidence, and Risk Attitude'. *Journal of Behavioral Decision Making* 17 (4), 297–311

Cancer Research Uk (n.d.) [online] available from <https://www.cancerresearchuk.org/health-professional/cancer-statistics/incidence/common-cancers-compared> [12 September 2020]

Carney, P.A., Bogart, A., Sickles, E.A., Smith, R., Buist, D.S.M., Kerlikowske, K., Onega, T., Miglioretti, D.L., Rosenberg, R., Yankaskas, B.C., and Geller, B.M. (2013) 'Feasibility and Acceptability of Conducting a Randomized Clinical Trial Designed to Improve Interpretation of Screening Mammography'. *Academic Radiology* 20 (11), 1389–1398

Carney, P.A., Miglioretti, D.L., Yankaskas, B.C., Kerlikowske, K., Rosenberg, R., Rutter, C.M., Geller, B.M., Abraham, L.A., Taplin, S.H., Dignan, M., Cutter, G., and Ballard-Barbash, R. (2003) 'Individual and Combined Effects of Age, Breast Density, and Hormone Replacement Therapy Use on the Accuracy of Screening Mammography.' *Annals of Internal Medicine* 138 (3), 168–75

Castaneda, C., Nalley, K., Mannion, C., Bhattacharyya, P., Blake, P., Pecora, A., Goy, A., and Suh, K.S. (2015) 'Clinical Decision Support Systems for Improving Diagnostic Accuracy and Achieving Precision Medicine'. *Journal of Clinical Bioinformatics* 5 (1), 4

Catchpole, K. (2013) 'Toward the Modelling of Safety Violations in Healthcare Systems.' *BMJ Quality & Safety* 22 (9), 705–9

Chawla-Duggan, R. (2007) 'Breaking out, Breaking through: Accessing Knowledge in a Non-western Overseas Educational Setting—Methodological Issues for an Outsider'. *Compare: A Journal of Comparative and International Education* 37 (2), 185–200

Chilamkurthy, S., Ghosh, R., Tanamala, S., Biviji, M., Campeau, N.G., Venugopal, V.K., Mahajan, V., Rao, P., and Warier, P. (2018) 'Deep Learning Algorithms for Detection of Critical Findings in Head CT Scans: A Retrospective Study'. *The Lancet* 392 (10162), 2388–2396

Christensen, C., Heckerung, P., Mackesy-Amiti, M.E., Bernstein, L.M., and Elstein, A.S. (1995) 'Pervasiveness of Framing Effects among Physicians and Medical Students'. *Journal of Behavioral Decision Making* 8 (3), 169–180

Christensen, C., Larson, J.R., Abbott, A., Ardolino, A., Franz, T., and Pfeiffer, C. (2000) 'Decision Making of Clinical Teams': *Medical Decision Making* 20 (1), 45–50

Ciatto, S., Ambrogetti, D., Bonardi, R., Catarzi, S., Risso, G., Rosselli Del Turco, M., and Mantellini, P. (2005) 'Second Reading of Screening Mammograms Increases Cancer Detection and Recall Rates. Results in the Florence Screening Programme'. *Journal of Medical Screening* 12 (2), 103–106

Ciatto, S., Houssami, N., Bernardi, D., Caumo, F., Pellegrini, M., Brunelli, S.,

Tuttobene, P., Bricolo, P., Fantò, C., Valentini, M., Montemezzi, S., and Macaskill, P. (2013) 'Integration of 3D Digital Mammography with Tomosynthesis for Population Breast-Cancer Screening (STORM): A Prospective Comparison Study'. *The Lancet Oncology* 14 (7), 583–589

Clarke, V. and Braun, V. (2013) *Successful Qualitative Research : A Practical Guide for Beginners*. 1st edn. SAGE

Coderre, S., Mandin, H., Harasym, P.H., and Fick, G.H. (2003) 'Diagnostic Reasoning Strategies and Diagnostic Success.' *Medical Education* 37 (8), 695–703

Colvin, C.J., de Heer, J., Winterton, L., Mellenkamp, M., Glenton, C., Noyes, J., Lewin, S., and Rashidian, A. (2013) 'A Systematic Review of Qualitative Evidence on Barriers and Facilitators to the Implementation of Task-Shifting in Midwifery Services'. *Midwifery* 29 (10), 1211–1221

Conant, E.F., Barlow, W.E., Herschorn, S.D., Weaver, D.L., Beaber, E.F., Tosteson, A.N.A., Haas, J.S., Lowry, K.P., Stout, N.K., Trentham-Dietz, A., Diflorio-Alexander, R.M., Li, C.I., Schnall, M.D., Onega, T., and Sprague, B.L. (2019) 'Association of Digital Breast Tomosynthesis vs Digital Mammography with Cancer Detection and Recall Rates by Age and Breast Density'. *JAMA Oncology* 5 (5), 635–642

Cook, R.D. and Weisberg, S. (1982) *Residuals and Influence in Regression*. Chapman and Hall

Coolican, H. (2004) *Research Methods and Statistics in Psychology*. Hodder & Stoughton

Coombs, M. and Ersser, S.J. (2004) 'Medical Hegemony in Decision-Making - a Barrier to Interdisciplinary Working in Intensive Care?' *Journal of Advanced Nursing* 46 (3), 245–252

Cornford, E., Reed, J., Murphy, A., Bennett, R., and Evans, A. (2011) 'Optimal Screening Mammography Reading Volumes; Evidence from Real Life in the East Midlands Region of the NHS Breast Screening Programme.' *Clinical Radiology* 66 (2), 103–7

Cornford, E.J., Evans, A.J., James, J.J., Burrell, H.C., Pinder, S.E., and Wilson, A.R.M. (2005) 'The Pathological and Radiological Features of Screen-Detected Breast Cancers Diagnosed Following Arbitration of Discordant Double Reading Opinions.' *Clinical Radiology* 60 (11), 1182–7

Crebbin, W., Beasley, S.W., and Watters, D.A.K. (2013) 'Clinical Decision Making: How Surgeons Do It'. *ANZ Journal of Surgery* 83 (6), 422–428

Cresswell, K.M., Panesar, S.S., Salvilla, S.A., Carson-Stevens, A., Larizgoitia, I., Donaldson, L.J., Bates, D., Sheikh, A., and Group, on behalf of the W.H.O. (WHO) S.P.C.E.W. (2013) 'Global Research Priorities to Better Understand the Burden of Iatrogenic Harm in Primary Care: An International Delphi Exercise'. *PLoS Medicine* 10 (11), e1001554

Creswell, J.W. and Plano Clark, V.L. (2011) *Designing and Conducting Mixed Methods Research*. SAGE Publications

Croskerry, P. (2012) 'Perspectives on Diagnostic Failure and Patient Safety.' *Healthcare Quarterly (Toronto, Ont.)* 15 Spec No, 50–6

Croskerry, P. (2009) 'A Universal Model of Diagnostic Reasoning'. *Academic Medicine* 84 (8), 1022–1028

Croskerry, P. (2003) 'The Importance of Cognitive Errors in Diagnosis and Strategies to Minimize Them.' *Academic Medicine : Journal of the Association of American Medical Colleges* 78 (8), 775–80

Croskerry, P. and Nimmo, G. (2011) 'Better Clinical Decision Making and Reducing Diagnostic Error'. *The Journal of the Royal College of Physicians of Edinburgh* 41 (2), 155–162

Croskerry, P., Singhal, G., and Mamede, S. (2013) 'Cognitive Debiasing 1: Origins of Bias and Theory of Debiasing'. *BMJ Quality & Safety* 22 (Suppl 2), ii58–ii64

Crotty, M. (1998) *The Foundations of Social Research : Meaning and Perspective in the Research Process*. Sage Publications

Cruz, D.M., Pimenta, C.M., and Lunney, M. (2009) 'Improving Critical Thinking and Clinical Reasoning with a Continuing Education Course'. *Journal of Continuing Education in Nursing* 40 (3), 121—127

Çubukcu, Z. (2006) *CRITICAL THINKING DISPOSITIONS OF THE TURKISH TEACHER CANDIDATES*. vol. 5

Culleton, B. (2015) 'Evidence-Based Decision-Making 4: Development and Limitations of Clinical Practice Guidelines.' *Methods in Molecular Biology (Clifton, N.J.)* 1281, 443–53

Culpan, A.M. (2016) 'Radiographer Involvement in Mammography Image Interpretation: A Survey of United Kingdom Practice'. *Radiography* 22 (4), 306–312

Curry, L. (Leslie A.. and Nunez-Smith, M. (2015) *Mixed Methods in Health Sciences Research : A Practical Primer*. 2455 Teller Road, Thousand Oaks California 91320: SAGE Publications Ltd

Curry, L.A., Brault, M.A., Linnander, E.L., McNatt, Z., Brewster, A.L., Cherlin, E., Flieger, S.P., Ting, H.H., and Bradley, E.H. (2018) 'Influencing Organisational Culture to Improve Hospital Performance in Care of Patients with Acute Myocardial Infarction: A Mixed-Methods Intervention Study.' *BMJ Quality & Safety* 27 (3), 207–217

Custers, E.J., Regehr, G., and Norman, G.R. (1996) 'Mental Representations of Medical Diagnostic Knowledge: A Review.' *Academic Medicine : Journal of the Association of American Medical Colleges* 71 (10 Suppl), S55-61

D'Orsi, C.J. (2013) *ACR BI-RADS Atlas: Breast Imaging Reporting and Data System*.

Reston, VA, American College of Radiology

D'orsi, C.J. and Sickles, E.A. (2017) '2017 Breast Caner Surveillance Consortium Reports on Interpretive Performance at Screening and Diagnostic Mammography: Welcome New Data, But Not as Benchmarks for Practice'. *Reviews and CommentaRy n Editorial Radiology* 283, 7–9

Davis, J.H. (1992) 'Some Compelling Intuitions about Group Consensus Decisions, Theoretical and Empirical Research, and Interpersonal Aggregation Phenomena: Selected Examples 1950–1990'. *Organizational Behavior and Human Decision Processes* 52 (1), 3–38

Dawson, N. V and Arkes, H.R. (1987) 'Systematic Errors in Medical Decision Making: Judgment Limitations.' *Journal of General Internal Medicine* 2 (3), 183–7

Deary, I.J. (2000) *Looking Down on Human Intelligence: From Psychometrics to the Brain*. New York: Oxford University Press

Debono, J.C., Poulos, A.E., Houssami, N., Turner, R.M., and Boyages, J. (2015) 'Evaluation of Radiographers' Mammography Screen-Reading Accuracy in Australia.' *Journal of Medical Radiation Sciences* 62 (1), 15–22

Denscombe, M. (2017) *The Good Research Guide : For Small-Scale Social Research Projects*. 6th edn. Open University Press

Denscombe, M. (2008) 'Communities of Practice'. *Journal of Mixed Methods*

Research 2 (3), 270–283

Destounis, S., Johnston, L., Highnam, R., Arieno, A., Morgan, R., and Chan, A. (2017) 'Using Volumetric Breast Density to Quantify the Potential Masking Risk of Mammographic Density'. *American Journal of Roentgenology* 208 (1), 222–227

Devine, D.J. and Philips, J.L. (2001) 'Do Smarter Teams Do Better: A Meta-Analysis of Cognitive Ability and Team Performance'. *Small Group Research* 32 (5), 507–532

Diamond, M.A. and Allcorn, S. (2009) 'Silo Mentality'. in *Private Selves in Public Organizations*. New York: Palgrave Macmillan US, 49–72

Diefenbach, M.A., Weinstein, N.D., and O'Reilly, J. (1993) 'Scales for Assessing Perceptions of Health Hazard Susceptibility'. *Health Education Research* 8 (2), 181–192

Dietrich, H. and Ehrlenspiel, F. (2010) 'Cognitive Interviewing: A Qualitative Tool for Improving Questionnaires in Sport Science'. *Measurement in Physical Education and Exercise Science* 14 (1), 51–60

Dinnes, J., Moss, S., Melia, J., Blanks, R., Song, F., and Kleijnen, J. (2001) 'Effectiveness and Cost-Effectiveness of Double Reading of Mammograms in Breast Cancer Screening: Findings of a Systematic Review.' *Breast (Edinburgh, Scotland)* 10 (6), 455–63

Dolan, J.G. (1999) 'A Method for Evaluating Health Care Providers' Decision Making'. *Medical Decision Making* 19 (1), 38–41

Donald, J.J. and Barnard, S.A. (2012) 'Common Patterns in 558 Diagnostic Radiology Errors'. *Journal of Medical Imaging and Radiation Oncology* 56 (2), 173–178

Draper, H., Wilson, S., Flanagan, S., and Ives, J. (2009) 'Offering Payments, Reimbursement and Incentives to Patients and Family Doctors to Encourage Participation in Research'. *Family Practice* 26 (3), 231–238

Duffy, S.W. and Gabe, R. (2005) 'What Should the Detection Rates of Cancers Be in Breast Screening Programmes?' *British Journal of Cancer* 92 (3), 597–600

Duffy, S.W., Nagtegaal, I.D., Astley, S.M., Gillan, M.G., McGee, M.A., Boggis, C.R., Wilson, M., Beetles, U.M., Griffiths, M.A., Jain, A.K., Johnson, J., Roberts, R., Deans, H., Duncan, K.A., Iyengar, G., Griffiths, P.M., Warwick, J., Cuzick, J., and Gilbert, F.J. (2008) 'Visually Assessed Breast Density, Breast Cancer Risk and the Importance of the Craniocaudal View'. *Breast Cancer Research* 10 (4), R64

Duijm, L.E.M., Groenewoud, J.H., Hendriks, J.H.C.L., and de Koning, H.J. (2004) 'Independent Double Reading of Screening Mammograms in the Netherlands: Effect of Arbitration Following Reader Disagreements'. *Radiology* 231 (2), 564–570

Duijm, L.E.M., Louwman, M.W.J., Groenewoud, J.H., van de Poll-Franse, L. V, Fracheboud, J., and Coebergh, J.W. (2009) 'Inter-Observer Variability in

Mammography Screening and Effect of Type and Number of Readers on Screening Outcome.' *British Journal of Cancer* 100 (6), 901–7

Duncan, K.A. and Scott, N.W. (2011) 'Is Film-Reading Performance Related to the Number of Films Read? The Scottish Experience.' *Clinical Radiology* 66 (2), 99–102

Durand, M.A., Haas, B.M., Yao, X., Geisel, J.L., Raghu, M., Hooley, R.J., Horvath, L.J., and Philpotts, L.E. (2015) 'Early Clinical Experience with Digital Breast Tomosynthesis for Screening Mammography.' *Radiology* 274 (1), 85–92

Eddy, D.M. (1990) 'Practice Policies: Where Do They Come From?' *JAMA* 263 (9), 1265–1275

Edmondson, A. C., R. M. Bohmer, and G.P.P. (2001) 'Speeding Up Team Learning'. *Harvard Business Review* 79 (9), 125–32

Edwards, P. (2010) 'Questionnaires in Clinical Trials: Guidelines for Optimal Design and Administration'. *Trials* 11 (2), 1–8

Einhorn, H.J., Hogarth, R.M., and Klempner, E. (1977) 'Quality of Group Judgment.' *Psychological Bulletin* 84 (1), 158–172

El-Kareh, R., Hasan, O., and Schiff, G.D. (2013) 'Use of Health Information Technology to Reduce Diagnostic Errors'. *BMJ Quality & Safety* 22 (Suppl 2), ii40

Elliott, N. (2010) '"Mutual Intacting": A Grounded Theory Study of Clinical

Judgement Practice Issues'. *Journal of Advanced Nursing* 66 (12), 2711–2721

Ellis, P. (2016) *Understanding Research for Nursing Students*. 3rd edn. SAGE Publications Ltd

Elmore, J.G., Jackson, S.L., Abraham, L., Miglioretti, D.L., Carney, P.A., Geller, B.M., Yankaskas, B.C., Kerlikowske, K., Onega, T., Rosenberg, R.D., Sickles, E.A., and Buist, D.S.M. (2009) 'Variability in Interpretive Performance at Screening Mammography and Radiologists' Characteristics Associated with Accuracy'. *Radiology* 253 (3), 641–651

Elmore, J.G., Nakano, C.Y., Koepsell, T.D., Desnick, L.M., D'Orsi, C.J., and Ransohoff, D.F. (2003) 'International Variation in Screening Mammography Interpretations in Community-Based Programs'. *JNCI Journal of the National Cancer Institute* 95 (18), 1384–1393

Elstein, A.S., Schwartz, A., and Schwarz, A. (2002) 'Clinical Problem Solving and Diagnostic Decision Making: Selective Review of the Cognitive Literature.' *BMJ (Clinical Research Ed.)* 324 (7339), 729–32

Eng, A., Gallant, Z., Shepherd, J., McCormack, V., Li, J., Dowsett, M., Vinnicombe, S., Allen, S., and Dos-Santos-Silva, I. (2014) 'Digital Mammographic Density and Breast Cancer Risk: A Case–Control Study of Six Alternative Density Assessment Methods'. *Breast Cancer Research* 16 (5), 439–451

Engmann, N.J., Golmakani, M.K., Miglioretti, D.L., Sprague, B.L., Kerlikowske, K., and

Breast Cancer Surveillance Consortium (2017) 'Population-Attributable Risk Proportion of Clinical Risk Factors for Breast Cancer'. *JAMA Oncology* 3 (9), 1228–1236

Erasmus, E., Gilson, L., Govender, V., and Nkosi, M. (2017) 'Organisational Culture and Trust as Influences over the Implementation of Equity-Oriented Policy in Two South African Case Study Hospitals Lucy Gilson'. *International Journal for Equity in Health* 16 (1), 1–14

Ericsson, K.A. (2004) 'Deliberate Practice and the Acquisition and Maintenance of Expert Performance in Medicine and Related Domains'. *Academic Medicine* 79 (10 SUPPL.), 70–81

Erzberger, C. and Kelle, U. (2003) *Making Inferences in Mixed Methods: The Rules of Integration in A. Tashakkori, & C. Teddlie (Eds.), Handbook of Mixed Methods in Social and Behavioral Research: 457-488*. Thousand Oaks, CA: Sage

Esteva, A., Kuprel, B., Novoa, R.A., Ko, J., Swetter, S.M., Blau, H.M., and Thrun, S. (2017) 'Dermatologist-Level Classification of Skin Cancer with Deep Neural Networks'. *Nature* 542 (7639), 115–118

Euler-Chelpin, M. von, Lillholm, M., Napolitano, G., Vejborg, I., Nielsen, M., and Lynge, E. (2018) 'Screening Mammography: Benefit of Double Reading by Breast Density.' *Breast Cancer Research and Treatment* 171 (3), 767–776

European Commission Initiative on Breast Cancer. Recommendations from European

Breast Guidelines. 3.0.0 Ed (2019)

Evans, D.G., Astley, S., Stavrinos, P., Harkness, E., Donnelly, L.S., Dawe, S., Jacob, I., Harvie, M., Cuzick, J., Brentnall, A., Wilson, M., Harrison, F., Payne, K., and Howell, A. (2016) 'Improvement in Risk Prediction, Early Detection and Prevention of Breast Cancer in the NHS Breast Screening Programme and Family History Clinics: A Dual Cohort Study.' *Programme Grants for Applied Research* 4 (11), 1–210

Evans, J.S.B.T. (2003) 'In Two Minds: Dual-Process Accounts of Reasoning.' *Trends in Cognitive Sciences* 7 (10), 454–9

Evans, J.S.B.T. and Stanovich, K.E. (2013) 'Dual-Process Theories of Higher Cognition'. *Perspectives on Psychological Science* 8 (3), 223–241

Everest, G., Fitzgerald, J., and Tate, L. (2014) 'Implementing NHS Culture Change: Contributions from Occupational Psychology'. *Clinical Psychology Forum* 2014 (263), 43–45

Faraway, J.J. (2014) *Linear Models with R*. 2nd edn. CRC Press

Fenwick, T., Seville, E., and Brunsdon, D. (2009) *Reducing the Impact of Organisational Silos on Resilience: A Report on the Impact of Silos on Resilience and How the Impacts Might Be Reduced*. held 2009

Field, A.P. (2009) *Discovering Statistics Using SPSS : (And Sex and Drugs and Rock 'n'*

Roll). 3rd edn. SAGE Publications

Finefter-Rosenbluh, I. (2017) 'Incorporating Perspective Taking in Reflexivity'. *International Journal of Qualitative Methods* 16 (1), 1–11

Fink, A. (2013) *How to Conduct Surveys: A Step-by-Step Guide*. Sage PublicationsSage CA: Thousand Oaks, CA

Fleming, M., Smith, S., Slaunwhite, J., and Sullivan, J. (2006) 'Investigating Interpersonal Competencies of Cardiac Surgery Teams.' *Canadian Journal of Surgery. Journal Canadien de Chirurgie* 49 (1), 22–30

Flin, R., Yule, S., McKenzie, L., Paterson-Brown, S., and Maran, N. (2006) 'Attitudes to Teamwork and Safety in the Operating Theatre.' *The Surgeon : Journal of the Royal Colleges of Surgeons of Edinburgh and Ireland* 4 (3), 145–51

Fong, Y., Evans, J., Brookes, D., and Thomas, K.G. (2012) 'Classification of Interval Breast Cancers and Their Presentation since Screening: Impact on Long-Term Survival'. *Breast Cancer Research* 14 (S1), P47

Ford, D., Easton, D.F., Stratton, M., Narod, S., Goldgar, D., Devilee, P., Bishop, D.T., Weber, B., Lenoir, G., Chang-Claude, J., Sobol, H., Teare, M.D., Struewing, J., Arason, A., Scherneck, S., Peto, J., Rebbeck, T.R., Tonin, P., Neuhausen, S., Barkardottir, R., Eyfjord, J., Lynch, H., Ponder, B.A., Gayther, S.A., and Zelada-Hedman, M. (1998) 'Genetic Heterogeneity and Penetrance Analysis of the BRCA1 and BRCA2 Genes in Breast Cancer Families. The Breast Cancer Linkage

Consortium.' *American Journal of Human Genetics* 62 (3), 676–89

Forrest, P. and Great Britain. Department of Health and Social Security. (1986) *Breast Cancer Screening : Report to the Health Ministers of England, Wales, Scotland & Northern Ireland*. H.M.S.O

Fox, J. (1991) *Regression Diagnostics*. 2455 Teller Road, Thousand Oaks California 91320 United States of America: SAGE Publications, Inc.

Francis, A., Thomas, J., Fallowfield, L., Wallis, M., Bartlett, J.M.S., Brookes, C., Roberts, T., Pirrie, S., Gaunt, C., Young, J., Billingham, L., Dodwell, D., Hanby, A., Pinder, S.E., Evans, A., Reed, M., Jenkins, V., Matthews, L., Wilcox, M., Fairbrother, P., Bowden, S., and Rea, D. (2015) 'Addressing Overtreatment of Screen Detected DCIS; the LORIS Trial'. *European Journal of Cancer* 51 (16), 2296–2303

Friedewald, S.M., Rafferty, E.A., Rose, S.L., Durand, M.A., Plecha, D.M., Greenberg, J.S., Hayes, M.K., Copit, D.S., Carlson, K.L., Cink, T.M., Barke, L.D., Greer, L.N., Miller, D.P., and Conant, E.F. (2014) 'Breast Cancer Screening Using Tomosynthesis in Combination With Digital Mammography'. *JAMA* 311 (24), 2499

Friedman, A. (2016) '"There Are Two People at Work That I'm Fairly Certain Are Black": Uncertainty and Deliberative Thinking in Blind Race Attribution'. *The Sociological Quarterly* 57 (3), 437–461

Friedman, C.P., Gatti, G.G., Franz, T.M., Murphy, G.C., Wolf, F.M., Heckerling, P.S., Fine, P.L., Miller, T.M., and Elstein, A.S. (2005) 'Do Physicians Know When Their Diagnoses Are Correct?' *Journal of General Internal Medicine* 20 (4), 334–339

FRQA (2017) *Film Reading QA Resource*.

Gale, A. (2010) 'Maintaining Quality in the UK Breast Screening Program'. *Medical Imaging 2010: Image Perception, Observer Performance, and Technology Assessment* 7627, 762702

Gardner, H.K., Gino, F., and Staats, B.R. (2012) 'Dynamically Integrating Knowledge in Teams: Transforming Resources into Performance'. *Academy of Management Journal* 55 (4), 998–1022

Garland, R. (1991) 'The Mid-Point on a Rating Scale: Is It Desirable?' *Marketing Bulletin* 2, 66–70

Gerrish, K, Lacey, A. (2012) *The Research Process in Nursing, 6th Edition | Nursing Research | Nursing General | Subjects | Wiley*. 6th edn. Wiley-Blackwell

Giess, C.S., Wang, A., Ip, I.K., Lacson, R., Pourjabbar, S., and Khorasani, R. (2019) 'Patient, Radiologist, and Examination Characteristics Affecting Screening Mammography Recall Rates in a Large Academic Practice'. *Journal of the American College of Radiology* 16 (4), 411–418

Gigerenzer, G., Hoffrage, U., and Kleinbölting, H. (1991) 'Probabilistic Mental

Models: A Brunswikian Theory of Confidence.' *Psychological Review* 98 (4), 506–528

Gilbert, F.J., Astley, S.M., Gillan, M.G.C., Agbaje, O.F., Wallis, M.G., James, J., Boggis, C.R.M., and Duffy, S.W. (2008) 'Single Reading with Computer-Aided Detection for Screening Mammography'. *New England Journal of Medicine* 359 (16), 1675–1684

Gilbert, F.J., Tucker, L., Gillan, M.G., Willsher, P., Cooke, J., Duncan, K.A., Michell, M.J., Dobson, H.M., Lim, Y.Y., Purushothaman, H., Strudley, C., Astley, S.M., Morrish, O., Young, K.C., and Duffy, S.W. (2015) 'The TOMMY Trial: A Comparison of TOMosynthesis with Digital MammographY in the UK NHS Breast Screening Programme – a Multicentre Retrospective Reading Study Comparing the Diagnostic Performance of Digital Breast Tomosynthesis and Digital Mammography With '. *Health Technology Assessment* 19 (4), 1–136

Gillies, R.J., Kinahan, P.E., and Hricak, H. (2016) 'Radiomics: Images Are More than Pictures, They Are Data'. *Radiology* 278 (2), 563–577

Ginsburg, L. and Gilin Oore, D. (2016) 'Patient Safety Climate Strength: A Concept That Requires More Attention'. *BMJ Quality &Amp; Safety* 25 (9), 680 LP – 687

Glass, G. V, Peckham, P.D., and Sanders, J.R. (1972) 'Consequences of Failure to Meet Assumptions Underlying the Fixed Effects Analyses of Variance and Covariance'. *Review of Educational Research* 42 (3), 237–288

Glogowska, M., Young, P., and Lockyer, L. (2011) 'Propriety, Process and Purpose: Considerations of the Use of the Telephone Interview Method in an Educational Research Study'. *Higher Education* 62 (1), 17–26

Goergen, S.K., Evans, J., Cohen, G.P., and MacMillan, J.H. (1997) 'Characteristics of Breast Carcinomas Missed by Screening Radiologists.' *Radiology* 204 (1), 131–135

Gøtzsche, P.C. (2012) 'Mammography Screening: Truth, Lies, and Controversy'. *The Lancet* 380 (9838), 218

Graber, M.L., Franklin, N., and Gordon, R. (2005) 'Diagnostic Error in Internal Medicine'. *Archives of Internal Medicine* 165 (13), 1493–1499

Grabler, P., Sighoko, D., Wang, L., Allgood, K., and Ansell, D. (2017) 'Recall and Cancer Detection Rates for Screening Mammography: Finding the Sweet Spot'. *American Journal of Roentgenology* 208 (1), 208–213

Gram, I.T., Funkhouser, E., and Tabár, L. (1997) 'The Tabár Classification of Mammographic Parenchymal Patterns.' *European Journal of Radiology* 24 (2), 131–136

Green, B., Oeppen, R.S., Smith, D.W., and Brennan, P.A. (2017) 'Challenging Hierarchy in Healthcare Teams – Ways to Flatten Gradients to Improve Teamwork and Patient Care'. *British Journal of Oral and Maxillofacial Surgery* 55 (5), 449–453

Greenhalgh, T. (2008) ‘How to Read a Paper : The Basics of Evidence-Based Medicine’. *BMJ* 336, 7657

Grofman, B., Feld, S.L., and Owen, G. (1984) ‘Group Size and the Performance of a Composite Group Majority: Statistical Truths and Empirical Results’. *Organizational Behavior and Human Performance* 33 (3), 350–359

Grumbach, K. and Bodenheimer, T. (2004) ‘Can Health Care Teams Improve Primary Care Practice?’ *JAMA* 291 (10), 1246–1251

Gubern-Mérida, A. (n.d.) ‘AI in Breast Imaging: Past, Present and (near) Future’. in *ESR Online 27 February-3 March 2019*.

Guest, G., Bunce, A., and Johnson, L. (2006) ‘How Many Interviews Are Enough?’ *Field Methods* 18 (1), 59–82

Gulshan, V., Peng, L., Coram, M., Stumpe, M.C., Wu, D., Narayanaswamy, A., Venugopalan, S., Widner, K., Madams, T., Cuadros, J., Kim, R., Raman, R., Nelson, P.C., Mega, J.L., and Webster, D.R. (2016) ‘Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs’. *JAMA* 316 (22), 2402

Gur, D., Sumkin, J.H., Hardesty, L.A., Clearfield, R.J., Cohen, C.S., Ganott, M.A., Hakim, C.M., Harris, K.M., Poller, W.R., Shah, R., Wallace, L.P., and Rockette, H.E. (2004) ‘Recall and Detection Rates in Screening Mammography’. *Cancer* 100 (8), 1590–1594

Haas, B.M., Kalra, V., Geisel, J., Raghu, M., Durand, M., and Philpotts, L.E. (2013) 'Comparison of Tomosynthesis Plus Digital Mammography and Digital Mammography Alone for Breast Cancer Screening'. *Radiology* 269 (3), 694–700

Hackney, L., Szczepura, A., Moody, L., and Whiteman, B. (2017) 'Review of the Evidence on the Use of Arbitration or Consensus within Breast Screening: A Systematic Scoping Review'. in *Radiography*.

Hagell, P., Hedin, P.J., Meads, D.M., Nyberg, L., and McKenna, S.P. (2010) 'Effects of Method of Translation of Patient-Reported Health Outcome Questionnaires: A Randomized Study of the Translation of the Rheumatoid Arthritis Quality of Life (RAQoL) Instrument for Sweden'. *Value in Health* 13 (4), 424–430

Handley, S.J. and Trippas, D. (2015) 'Dual Processes and the Interplay between Knowledge and Structure: A New Parallel Processing Model'. in *Psychology of Learning and Motivation - Advances in Research and Theory*. vol. 62. Academic Press Inc., 33–58

Hannagan, R.J. and Larimer, C.W. (2010) 'Does Gender Composition Affect Group Decision Outcomes? Evidence from a Laboratory Experiment'. *Political Behavior* 32 (1), 51–67

Hanoch, Y. and Wallin, A. (2003) 'Educating Intuition. R. M. Hogarth. University of Chicago Press, Chicago, 2001. No. of Pages 335. ISBN 0-226-34860-1. Price £19'. *Applied Cognitive Psychology* 17 (1), 122–124

Hardy, M., Snaith, B., and Scally, A. (2013) 'The Impact of Immediate Reporting on Interpretive Discrepancies and Patient Referral Pathways within the Emergency Department: A Randomised Controlled Trial.' *The British Journal of Radiology* 86 (1021), 20120112

Harvey Hugh, Heindl Andreas, Khara Galvin, Korkinof Dimitrios, O'Neill Michael, Yearsley Joseph, Karpati Edith, Rijken Tobias, Kecskemethy Peter, and F.G. (2019) 'Deep Learning in Breast Cancer Screening'. in *Artificial Intelligence in Medical Imaging,*. 1st edn. ed. by Ranschaert, E.R., Morozov, S., and Algra, P.R. Switzerland: Springer International Publishing, 187–215

Hatala, R., Norman, G.R., and Brooks, L.R. (1999) 'Impact of a Clinical Scenario on Accuracy of Electrocardiogram Interpretation.' *Journal of General Internal Medicine* 14 (2), 126–9

Hautz, W.E., Kämmer, J.E., Schauber, S.K., Spies, C.D., and Gaissmaier, W. (2015) 'Diagnostic Performance by Medical Students Working Individually or in Teams'. *JAMA* 313 (3), 303–304

Hayter, A.J. (1984) 'A Proof of the Conjecture That the Tukey-Kramer Multiple Comparisons Procedure Is Conservative'. *The Annals of Statistics* 12, 61–75

Heale, R. and Forbes, D. (2013) 'Understanding Triangulation in Research'. *Evidence-Based Nursing* 16 (4), 98

Helvie, M. (2007) 'Improving Mammographic Interpretation: Double Reading and

Computer-Aided Diagnosis'. *Radiologic Clinics of North America* 45 (5), 801–811

Heywang-Köbrunner, S.H., Hacker, A., and Sedlacek, S. (2011) 'Advantages and Disadvantages of Mammography Screening'. *Breast Care* 6 (3), 199–207

Higgs, J. (2008) *Clinical Reasoning in the Health Professions*. 3rd edn. BH/Elsevier

Hoff, S.R., Myklebust, T.-Å., Lee, C.I., and Hofvind, S. (2019) 'Influence of Mammography Volume on Radiologists' Performance: Results from BreastScreen Norway'. *Radiology* 182684

Hoff, S.R., Samset, J.H., Abrahamsen, A.-L., Vigeland, E., Klepp, O., and Hofvind, S. (2011) 'Missed and True Interval and Screen-Detected Breast Cancers in a Population Based Screening Program'. *Academic Radiology* 18 (4), 454–460

Hofvind, S., Geller, B.M., Rosenberg, R.D., and Skaane, P. (2009) 'Screening-Detected Breast Cancers: Discordant Independent Double Reading in a Population-Based Screening Program'. *Radiology* 253 (3), 652–660

Hofvind, S, Geller, B.M., Skelly, J., and Vacek, P.M. (2012) 'Sensitivity and Specificity of Mammographic Screening as Practised in Vermont and Norway'. *The British Journal of Radiology* 85 (1020), e1226–e1232

Hofvind, Solveig, Ponti, A., Patnick, J., Ascunce, N., Njor, S., Broeders, M., Giordano, L., Frigerio, A., and Törnberg, S. (2012) 'False-Positive Results in Mammographic Screening for Breast Cancer in Europe: A Literature Review and Survey of

Service Screening Programmes'. *Journal of Medical Screening* 19 (1_suppl), 57–66

Hofvind, S., Vacek, P.M., Skelly, J., Weaver, D.L., and Geller, B.M. (2008) 'Comparing Screening Mammography for Early Breast Cancer Detection in Vermont and Norway.' *Journal of the National Cancer Institute* 100 (15), 1082–91

Hogarth, R.M. (2010) *Educating Intuition*. 5th edn. University of Chicago Press

Holt, A. (2010) 'Using the Telephone for Narrative Interviewing: A Research Note'. *Qualitative Research* 10 (1), 113–121

Hope, R.A., Savulescu, J., and Hendrick, J. (2008) *Medical Ethics and Law : The Core Curriculum*. 2nd edn. Churchill Livingstone/Elsevier

Hopkins, C. (2011) *Quality Assurance Guidelines for Breast Cancer ScreeninG RadioloGy Second Edition* [online] available from <www.cancerscreening.nhs.uk> [4 October 2018]

Houssami, N., Lee, C.I., Buist, D.S.M., and Tao, D. (2017) 'Artificial Intelligence for Breast Cancer Screening: Opportunity or Hype?' *The Breast* 36, 31–33

Houssami, N. and Miglioretti, D.L. (2016) 'Digital Breast Tomosynthesis: A Brave New World of Mammography Screening.' *JAMA Oncology* 2 (6), 725–727

Hovda, T., Holen, Å.S., Lång, K., Albertsen, J.L., Bjørndal, H., Brandal, S.H.B., Sahlberg, K.K., Skaane, P., Suhrke, P., and Hofvind, S. (2020) 'Interval and Consecutive

Round Breast Cancer after Digital Breast Tomosynthesis and Synthetic 2d Mammography versus Standard 2d Digital Mammography in Breastscreen Norway'. *Radiology* 294 (2), 256–264

Howell, A., Sims, A.H., Ong, K.R., Harvie, M.N., Evans, D.G.R., and Clarke, R.B. (2005) 'Mechanisms of Disease: Prediction and Prevention of Breast Cancer—Cellular and Molecular Interactions'. *Nature Clinical Practice Oncology* 2 (12), 635–646

Huang, D.T., Clermont, G., Sexton, B.J., Karlo, C.A., Miller, R.G., Weissfeld, L.A., Rowan, K.M., and Angus, D.C. (2007) 'Perceptions of Safety Culture Vary across the Intensive Care Units of a Single Institution*'. *Critical Care Medicine* 35 (1), 165–176

Huber, P.J. and Ronchetti, E. (2009) *Robust Statistics*. 2nd edn. Wiley

Huczynski, A., Buchanan, D.A., Huczynski, A., and Buchanan, D.A. (n.d.) *Organizational Behaviour*.

Ilgen, J.S., Bowen, J.L., McIntyre, L.A., Banh, K. V., Barnes, D., Coates, W.C., Druck, J., Fix, M.L., Rimple, D., Yarris, L.M., and Eva, K.W. (2013) 'Comparing Diagnostic Performance and the Utility of Clinical Vignette-Based Assessment Under Testing Conditions Designed to Encourage Either Automatic or Analytic Thought'. *Academic Medicine* 88 (10), 1545–1551

Ilgen, J.S., Bowen, J.L., Yarris, L.M., Fu, R., Lowe, R.A., and Eva, K. (2011) 'Adjusting Our Lens: Can Developmental Differences in Diagnostic Reasoning Be

Harnessed to Improve Health Professional and Trainee Assessment?' *Academic Emergency Medicine* 18, S79–S86

Irvine, A., Drew, P., and Sainsbury, R. (2013) '"Am I Not Answering Your Questions Properly?" Clarification, Adequacy and Responsiveness in Semi-Structured Telephone and Face-to-Face Interviews'. *Qualitative Research* 13 (1), 87–106

Ivankova, N. and Kawamura, Y. (2010) 'Emerging Trends in the Utilization of Integrated Designs in the Social, Behavioral, and Health Sciences'. in *SAGE Handbook of Mixed Methods in Social & Behavioral Research*. 2455 Teller Road, Thousand Oaks California 91320 United States: SAGE Publications, Inc., 581–612

Jaccard, J. (1998) *Interaction Effects in Factorial Analysis of Variance*. 2455 Teller Road, Thousand Oaks California 91320 United States of America: SAGE Publications, Inc.

James, J.J. and Tennant, S.L. (2018) 'Contrast-Enhanced Spectral Mammography (CESM)'. *Clinical Radiology* 73 (8), 715–723

Jenkins, J., Murphy, A.E., Edmondson-Jones, M., Sibbering, D.M., and Turnbull, A.E. (2014) 'Film Reading in the East Midlands Breast Screening Programme -- Are We Missing Opportunities for Earlier Diagnosis?' *Clinical Radiology* 69 (4), 385–90

Jesmin, S., Thind, A., and Sarma, S. (2012) 'Does Team-Based Primary Health Care

Improve Patients' Perception of Outcomes? Evidence from the 2007–08 Canadian Survey of Experiences with Primary Health'. *Health Policy* 105 (1), 71–83

Johnson, A., Nguyen, H., Groth, M., Wang, K., and Ng, J.L. (2016) 'Time to Change: A Review of Organisational Culture Change in Health Care Organisations'. *Journal of Organizational Effectiveness: People and Performance* 3 (3), 265–288

Johnson, O.A. (2016) *AI Can Excel at Medical Diagnosis, but the Harder Task Is to Win Hearts and Minds First - Leeds Institute for Data Analytics*.

Johnson, T.R. and Edu, T.J. (2016) 'Violation of the Homogeneity of Regression Slopes Assumption in ANCOVA for Two-Group Pre-Post Designs: Tutorial on a Modied Johnson-Neyman Procedure'. *The Quantitative Methods for Psychology* 12 (3)

Jolley, J. (2013) *Introducing Research and Evidence-Based Practice for Nursing and Healthcare Professionals*. Pearson Education

Jones, T.L., Baxter, M.A.J., and Khanduja, V. (2013) 'A Quick Guide to Survey Research.' *Annals of the Royal College of Surgeons of England* 95 (1), 5–7

Joshi, I., Morley, J. (2019) *Artificial Intelligence: How to Get It Right*.

Juel, I.-M., Skaane, P., Hoff, S.R., Johannessen, G., and Hofvind, S. (2010) 'Screen-Film Mammography versus Full-Field Digital Mammography in a Population-

Based Screening Program: The Sogn and Fjordane Study'. *Acta Radiologica* 51 (9), 962–968

Kahneman, D. (2012) *Thinking, Fast and Slow*. Penguin

Kahneman, D. and Klein, G. (2009) 'Conditions for Intuitive Expertise: A Failure to Disagree.' *American Psychologist* 64 (6), 515–526

Kash, B.A., Cheon, O., Halzack, N.M., and Miller, T.R. (2018) 'Measuring Team Effectiveness in the Health Care Setting: An Inventory of Survey Tools'. *Health Services Insights* 11, 1178632918796230

Kattan, M.W., O'Rourke, C., Yu, C., and Chagin, K. (2016) 'The Wisdom of Crowds of Doctors'. *Medical Decision Making* 36 (4), 536–540

Kavanagh, A.M., Giles, G.G., Mitchell, H., and Cawson, J.N. (2000) 'The Sensitivity, Specificity, and Positive Predictive Value of Screening Mammography and Symptomatic Status'. *Journal of Medical Screening* 7 (2), 105–110

Kazmer, M.M. and Xie, B. (2008) 'QUALITATIVE INTERVIEWING IN INTERNET STUDIES: Playing with the Media, Playing with the Method'. *Information, Communication & Society* 11 (2), 257–278

Kee, F., Owen, T., and Leathem, R. (2004) 'Decision Making in a Multidisciplinary Cancer Team: Does Team Discussion Result in Better Quality Decisions?' *Medical Decision Making* 24 (6), 602–613

Keele, R. (2011) *Nursing Research and Evidence-Based Practice : Ten Steps to Success*. Jones & Bartlett Learning

Kelly, S.E. (2012) 'Qualitative Interviewing Techniques and Styles'. in *The SAGE Handbook of Qualitative Methods in Health Research*. 1 Oliver's Yard, 55 City Road, London EC1Y 1SP United Kingdom: SAGE Publications Ltd, 307–326

Khullar, D., Jha, A.K., and Jena, A.B. (2015) 'Reducing Diagnostic Errors--Why Now?' *The New England Journal of Medicine* 373 (26), 2491–3

Kirk, R. (2013) *Experimental Design: Procedures for the Behavioral Sciences*. 2455 Teller Road, Thousand Oaks California 91320 United States: SAGE Publications, Inc.

Klein, G.A. (1998) *Sources of Power : How People Make Decisions*. MIT Press

Klompenhouwer, E.G., Voogd, A.C., den Heeten, G.J., Strobbe, L.J.A., de Haan, A.F.J., Wauters, C.A., Broeders, M.J.M., and Duijm, L.E.M. (2015a) 'Blinded Double Reading Yields a Higher Programme Sensitivity than Non-Blinded Double Reading at Digital Screening Mammography: A Prospected Population Based Study in the South of The Netherlands'. *European Journal of Cancer* 51 (3), 391–399

Klompenhouwer, E.G., Voogd, A.C., den Heeten, G.J., Strobbe, L.J.A., Tjan-Heijnen, V.C., Broeders, M.J.M., and Duijm, L.E.M. (2015b) 'Discrepant Screening Mammography Assessments at Blinded and Non-Blinded Double Reading:

Impact of Arbitration by a Third Reader on Screening Outcome'. *European Radiology* 25 (10), 2821–2829

Kooi, T., Litjens, G., van Ginneken, B., Gubern-Mérida, A., Sánchez, C.I., Mann, R., den Heeten, A., and Karssemeijer, N. (2017) 'Large Scale Deep Learning for Computer Aided Detection of Mammographic Lesions'. *Medical Image Analysis* 35, 303–312

Kopans, D.B., Smith, R.A., and Duffy, S.W. (2011) 'Mammographic Screening and "Overdiagnosis"'. *Radiology* 260 (3), 616–620

Koriat, A. (2012) 'When Are Two Heads Better than One and Why?' *Science* 336 (6079), 360–362

Kozlowski, S.W.J. and Ilgen, D.R. (2006) 'Enhancing the Effectiveness of Work Groups and Teams'. *Psychological Science in the Public Interest* 7 (3), 77–124

Krosnick, J.A. and Presser, S. (2009) 'Question and Questionnaire Design'. in *Handbook of Survey Research*. 2nd edn. San Diego, CA: Elsevier

Kruglanski, A.W. and Gigerenzer, G. (2011) 'Intuitive and Deliberate Judgments Are Based on Common Principles.' *Psychological Review* 118 (1), 97–109

Kurvers, R.H.J.M., Herzog, S.M., Hertwig, R., Krause, J., Carney, P.A., Bogart, A., Argenziano, G., Zalaudek, I., and Wolf, M. (2016) 'Boosting Medical Diagnostics by Pooling Independent Judgments.' *Proceedings of the National Academy of*

Sciences of the United States of America 113 (31), 8777–82

Kurvers, R.H.J.M., Krause, J., Argenziano, G., Zalaudek, I., and Wolf, M. (2015) 'Detection Accuracy of Collective Intelligence Assessments for Skin Cancer Diagnosis'. *JAMA Dermatology* 151 (12), 1346–1353

Kvale, S. and Brinkmann, S. (2014) *InterViews : Learning the Craft of Qualitative Research Interviewing*. third. SAGE Publications, Inc.

Laming, D. and Warren, R. (2000) 'Improving the Detection of Cancer in the Screening of Mammograms'. *Journal of Medical Screening* 7 (1), 24–30

Lång, K, Dustler, M, Dahlblom, V, Andersson, I, Zackrisson, S. (2019) *Can Artificial Intelligence Identify Normal Mammograms in Screening?*

Lång, K., Nergården, M., Andersson, I., Rosso, A., and Zackrisson, S. (2016) 'False Positives in Breast Cancer Screening with One-View Breast Tomosynthesis: An Analysis of Findings Leading to Recall, Work-up and Biopsy Rates in the Malmö Breast Tomosynthesis Screening Trial'. *European Radiology* 26 (11), 3899–3907

Lau, F. and Kuziemsky, C. (2017) *Handbook of EHealth Evaluation: An Evidence-Based Approach*.

Lechuga, V.M. (2012) 'Exploring Culture from a Distance: The Utility of Telephone Interviews in Qualitative Research'. *International Journal of Qualitative Studies in Education* 25 (3), 251–268

Lehman, C.D. (2020) ‘Artificial Intelligence to Support Independent Assessment of Screening Mammograms—The Time Has Come’. *JAMA Oncology* E1-2

Lehman, C.D., Arao, R.F., Sprague, B.L., Lee, J.M., Buist, D.S.M., Kerlikowske, K., Henderson, L.M., Onega, T., Tosteson, A.N.A., Rauscher, G.H., and Miglioretti, D.L. (2017) ‘National Performance Benchmarks for Modern Screening Digital Mammography: Update from the Breast Cancer Surveillance Consortium’. *Radiology* 283 (1), 49–58

Lehman, C.D., Wellman, R.D., Buist, D.S.M., Kerlikowske, K., Tosteson, A.N.A., and Miglioretti, D.L. (2015) ‘Diagnostic Accuracy of Digital Screening Mammography With and Without Computer-Aided Detection.’ *JAMA Internal Medicine* 175 (11), 1828–1837

Leppink, J. (2018) ‘Analysis of Covariance (ANCOVA) vs. Moderated Regression (MODREG): Why the Interaction Matters’. *Health Professions Education*

Levitin, D.J. (2011) *Foundations of Cognitive Psychology : Core Readings*. 2nd edn. Pearson Education

Lighthall, G.K. and Vazquez-Guillamet, C. (2015) ‘Understanding Decision Making in Critical Care.’ *Clinical Medicine & Research* 13 (3–4), 156–68

Lincoln, Y.S. and Guba, E.G. (1989) ‘Ethics: The Failure of Positivist Science’. *The Review of Higher Education* 12 (3), 221–240

Lipshitz, R., Klein, G., Orasanu, J., and Salas, E. (2001) 'Focus Article: Taking Stock of Naturalistic Decision Making'. *Journal of Behavioral Decision Making J. Behav. Dec. Making* 14 (5), 331–352

De Lisle, J. (2011) 'THE BENEFITS AND CHALLENGES OF MIXING METHODS AND METHODOLOGIES: Lessons Learnt From Implementing Qualitatively Led Mixed Methods Research Designs in Trinidad and Tobago The Emergence of Mixed Methods Research'. *Caribbean Curriculum* 18, 87–120

Liston, J.C. and Dall, B.J.G. (2003) 'Can the NHS Breast Screening Programme Afford Not to Double Read Screening Mammograms?' *Clinical Radiology* 58 (6), 474–7

Lomas, J. (1993) 'Diffusion, Dissemination, and Implementation: Who Should Do What?' *Annals of the New York Academy of Sciences* 703, 226–35; discussion 235-7

Lowry, K.P., Trentham-Dietz, A., Schechter, C.B., Alagoz, O., Barlow, W.E., Burnside, E.S., Conant, E.F., Hampton, J.M., Huang, H., Kerlikowske, K., Lee, S.J., Miglioretti, D.L., Sprague, B.L., Tosteson, A.N.A., Yaffe, M.J., and Stout, N.K. (2020) 'Long-Term Outcomes and Cost-Effectiveness of Breast Cancer Screening With Digital Breast Tomosynthesis in the United States.' *Journal of the National Cancer Institute* 112 (6), 582–589

Lu, J., Zhang, G., and Ruan, D. (2007) 'Intelligent Multi-Criteria Fuzzy Group Decision-Making for Situation Assessments'. *Soft Computing* 12 (3), 289–299

Mahmoodi, A., Bang, D., Olsen, K., Zhao, Y.A., Shi, Z., Broberg, K., Safavi, S., Han, S., Ahmadabadi, M.N., Frith, C.D., Roepstorff, A., Rees, G., and Bahrami, B. (2015) 'Equality Bias Impairs Collective Decision-Making across Cultures'. *Proceedings of the National Academy of Sciences* 112 (12), 3835–3840

Maltby, J., Williams, G., McGarry, J., and Day, L. (2010) *Research Methods for Nursing and Healthcare*. Pearson Education

Malvezzi, M., Carioli, G., Bertuccio, P., Boffetta, P., Levi, F., La Vecchia, C., and Negri, E. (2019) 'European Cancer Mortality Predictions for the Year 2019 with Focus on Breast Cancer'. *Annals of Oncology* 30 (5), 781–787

Mamede, S., van Gog, T., van den Berge, K., Rikers, R.M.J.P., van Saase, J.L.C.M., van Guldener, C., and Schmidt, H.G. (2010a) 'Effect of Availability Bias and Reflective Reasoning on Diagnostic Accuracy Among Internal Medicine Residents'. *JAMA* 304 (11), 1198

Mamede, S., van Gog, T., van den Berge, K., van Saase, J.L.C.M., and Schmidt, H.G. (2014) 'Why Do Doctors Make Mistakes? A Study of the Role of Salient Distracting Clinical Features'. *Academic Medicine* 89 (1), 114–120

Mamede, S. and Schmidt, H.G. (2014) 'Reflection in Diagnostic Reasoning'. *Academic Medicine* 89 (7), 959–960

Mamede, S. and Schmidt, H.G. (2004) 'The Structure of Reflective Practice in Medicine'. *Medical Education* 38 (12), 1302–1308

Mamede, S., Schmidt, H.G., and Penaforte, J.C. (2008) 'Effects of Reflective Practice on the Accuracy of Medical Diagnoses'. *Medical Education* 42 (5), 468–475

Mamede, S., Schmidt, H.G., Rikers, R.M.J.P., Custers, E.J.F.M., Splinter, T.A.W., and van Saase, J.L.C.M. (2010b) 'Conscious Thought Beats Deliberation without Attention in Diagnostic Decision-Making: At Least When You Are an Expert'. *Psychological Research* 74 (6), 586–592

Mamede, S., Splinter, T.A.W., van Gog, T., Rikers, R.M.J.P., and Schmidt, H.G. (2012) 'Exploring the Role of Salient Distracting Clinical Features in the Emergence of Diagnostic Errors and the Mechanisms through Which Reflection Counteracts Mistakes.' *BMJ Quality & Safety* 21 (4), 295–300

'Mammography Quality Standards Act' (1992) *USC* 42, 102–539

Mann, L., Radford, M., Burnett, P., Ford, S., Bond, M., Leung, K., Nakamura, H., Vaughan, G., and Yang, K. (1998) 'Cross-Cultural Differences in Self-Reported Decision-Making Style and Confidence'. *International Journal of Psychology* 33 (5), 325–335

MANSER, T. (2009) 'Teamwork and Patient Safety in Dynamic Domains of Healthcare: A Review of the Literature'. *Acta Anaesthesiologica Scandinavica* 53 (2), 143–151

Marinovich, M.L., Hunter, K.E., Macaskill, P., and Houssami, N. (2018) 'Breast Cancer Screening Using Tomosynthesis or Mammography: A Meta-Analysis of Cancer

Detection and Recall.' *Journal of the National Cancer Institute* 110 (9), 942–949

Mark, T.L., Johnson, G., Fortner, B., and Ryan, K. (2008) 'The Benefits and Challenges of Using Computer-Assisted Symptom Assessments in Oncology Clinics: Results of a Qualitative Assessment'. *Technology in Cancer Research & Treatment* 7 (5), 401–405

Marmot, M.G., Altman, D.G., Cameron, D.A., Dewar, J.A., Thompson, S.G., Wilcox, M., and Screening, T.I.U.P. on B.C. (2013) 'The Benefits and Harms of Breast Cancer Screening: An Independent Review.' *British Journal of Cancer* 108 (11), 2205–40

Mason, J. (2017) *Qualitative Researching*. Third. SAGE Publications Ltd

Massat, N.J., Douglas, E., Waller, J., Wardle, J., and Duffy, S.W. (2015) 'Variation in Cervical and Breast Cancer Screening Coverage in England: A Cross-Sectional Analysis to Characterise Districts with Atypical Behaviour.' *BMJ Open* 5 (7), e007735

Maxwell, A.J., Michell, M., Lim, Y.Y., Astley, S.M., Wilson, M., Hurley, E., Evans, D.G., Howell, A., Iqbal, A., Kotre, J., Duffy, S., and Morris, J. (2017) 'A Randomised Trial of Screening with Digital Breast Tomosynthesis plus Conventional Digital 2D Mammography versus 2D Mammography Alone in Younger Higher Risk Women'. *European Journal of Radiology* 94, 133–139

Maxwell, S.E., Delaney, H.D., and Kelley, K. (Professor of information technology)

(2017) *Designing Experiments and Analyzing Data : A Model Comparison Perspective*. 3rd edn. Routledge

McKinney, S.M., Sieniek, M., Godbole, V., Godwin, J., Antropova, N., Ashrafian, H., Back, T., Chesus, M., Corrado, G.C., Darzi, A., Etemadi, M., Garcia-Vicente, F., Gilbert, F.J., Halling-Brown, M., Hassabis, D., Jansen, S., Karthikesalingam, A., Kelly, C.J., King, D., Ledsam, J.R., Melnick, D., Mostofi, H., Peng, L., Reicher, J.J., Romera-Paredes, B., Sidebottom, R., Suleyman, M., Tse, D., Young, K.C., De Fauw, J., and Shetty, S. (2020) 'International Evaluation of an AI System for Breast Cancer Screening'. *Nature* 577 (7788), 89–94

McMurray, A. (1992) 'Expertise in Community Health Nursing'. *Journal of Community Health Nursing* 9 (2), 65–75

Mertens, D.M. (2005) *Research and Evaluation in Education and Psychology : Integrating Diversity with Quantitative, Qualitative, and Mixed Methods*. Sage Publications

Michell, M.J. and Batohi, B. (2018) 'Role of Tomosynthesis in Breast Imaging Going Forward.' *Clinical Radiology* 73 (4), 358–371

Miglioretti, D.L., Gard, C.C., Carney, P.A., Onega, T.L., Buist, D.S.M., Sickles, E.A., Kerlikowske, K., Rosenberg, R.D., Yankaskas, B.C., Geller, B.M., and Elmore, J.G. (2009) 'When Radiologists Perform Best: The Learning Curve in Screening Mammogram Interpretation.' *Radiology* 253 (3), 632–40

Miglioretti, D.L., Smith-Bindman, R., Abraham, L., Brenner, R.J., Carney, P.A., Bowles, E.J.A., Buist, D.S.M., and Elmore, J.G. (2007) 'Radiologist Characteristics Associated With Interpretive Performance of Diagnostic Mammography'. *JNCI: Journal of the National Cancer Institute* 99 (24), 1854–1863

Minda, J.P. and Smith, J.D. (2001) 'Prototypes in Category Learning: The Effects of Category Size, Category Structure, and Stimulus Complexity.' *Journal of Experimental Psychology. Learning, Memory, and Cognition* 27 (3), 775–99

Moacdieh, N. and Sarter, N. (2015) 'Display Clutter'. *Human Factors: The Journal of the Human Factors and Ergonomics Society* 57 (1), 61–100

Mohd Norsuddin, N., Reed, W., Mello-Thoms, C., and Lewis, S.J. (2015) 'Understanding Recall Rates in Screening Mammography: A Conceptual Framework Review of the Literature'. *Radiography* 21 (4), 334–341

Molina-Azorin, J.F. (2016) 'Mixed Methods Research: An Opportunity to Improve Our Studies and Our Research Skills'. *European Journal of Management and Business Economics* 25, 37–38

Monteiro, Sandra D, Sherbino, J., Patel, A., Mazzetti, I., Norman, G.R., and Howey, E. (2015) 'Reflecting on Diagnostic Errors: Taking a Second Look Is Not Enough.' *Journal of General Internal Medicine* 30 (9), 1270–4

Monteiro, Sandra D., Sherbino, J.D., Ilgen, J.S., Dore, K.L., Wood, T.J., Young, M.E., Bandiera, G., Blouin, D., Gaissmaier, W., Norman, G.R., and Howey, E. (2015)

'Disrupting Diagnostic Reasoning'. *Academic Medicine* 90 (4), 511–517

Montero, J. (2007) 'The Impact of Fuzziness in Social Choice Paradoxes'. *Soft Computing* 12 (2), 177–182

Moore, R. (2017) 'Fast or Slow: Sociological Implications of Measuring Dual-Process Cognition'. *Sociological Science* 4, 196–223

Moorhead, G. (1982) 'Groupthink: Hypothesis in Need of Testing'. *Group & Organization Studies* 7 (4), 429–444

Moran, S. and Warren-Forward, H. (2016) 'The Diagnostic Accuracy of Radiographers Assessing Screening Mammograms: A Systematic Review'. *Radiography* 22 (2), 137–146

Morgan, D.L. (1998) 'Practical Strategies for Combining Qualitative and Quantitative Methods: Applications to Health Research'. *Qualitative Health Research* 8 (3), 362–376

Morse, J., Swanson, J., and Kuzel, A. (2001) *The Nature of Qualitative Evidence*. 2455 Teller Road, Thousand Oaks California 91320 United States of America: SAGE Publications, Inc.

Moser, K., Sellars, S., Wheaton, M., Cooke, J., Duncan, A., Maxwell, A., Michell, M., Wilson, M., Beral, V., Peto, R., Richards, M., and Patnick, J. (2011) 'Extending the Age Range for Breast Screening in England: Pilot Study to Assess the

Feasibility and Acceptability of Randomization'. *Journal of Medical Screening* 18 (2), 96–102

Moule, P., Aveyard, H., and Goodman, M.L. (2016) *Nursing Research : An Introduction*. 3rd edn. SAGE Publications Ltd

Mullen, L.A., Panigrahi, B.B., Hollada, J., Panigrahi, B.B., Falomo, E.T., and Harvey, S.C. (2017) 'Strategies for Decreasing Screening Mammography Recall Rates While Maintaining Performance Metrics'. *Academic Radiology* 24 (12), 1556–1560

Muntanyola Saura, D. and Romero Balsas, P. (2014) 'Interviewing and Surveying over the Phone: A Reflexive Account of a Research on Parenting'. *Quality & Quantity* 48 (5), 2615–2630

Murphy, F.J. and Yielder, J. (2010) 'Establishing Rigour in Qualitative Radiography Research'. *Radiography* 16 (1), 62–67

Musselwhite, K., Cuff, L., McGregor, L., and King, K.M. (2007) 'The Telephone Interview Is an Effective Method of Data Collection in Clinical Nursing Research: A Discussion Paper'. *International Journal of Nursing Studies* 44 (6), 1064–1070

Nagpal, K., Vats, A., Lamb, B., Ashrafian, H., Sevdalis, N., Vincent, C., and Moorthy, K. (2010) 'Information Transfer and Communication in Surgery'. *Annals of Surgery* 252 (2), 225–239

National Breast Imaging Academy (n.d.) available from

<https://nationalbreastimagingacademy.org/> [2 August 2019]

National Cancer Institute (2020) *TMIST (Tomosynthesis Mammographic Imaging Screening Trial)* [online] available from <https://www.cancer.gov/about-cancer/treatment/clinical-trials/nci-supported/tmist> [7 September 2020]

Nelson, H.D., Pappas, M., Cantor, A., Griffin, J., Daeges, M., and Humphrey, L. (2016) 'Harms of Breast Cancer Screening: Systematic Review to Update the 2009 U.S. Preventive Services Task Force Recommendation'. *Annals of Internal Medicine* 164 (4), 256

NHS (2019) *The NHS Long Term Plan* [online] available from <https://www.longtermplan.nhs.uk/> [2 August 2019]

NHS (2014) *FIVE YEAR FORWARD VIEW* [online] available from <https://www.england.nhs.uk/five-year-forward-view/> [9 July 2019]

NHS England (2019) *Transforming Imaging Services in England: A National Strategy for Imaging Networks*. (November), 1–56

Noble, H. and Smith, J. (2015) 'Issues of Validity and Reliability in Qualitative Research.' *Evidence-Based Nursing* 18 (2), 34–5

Norman, G. (2005) 'Research in Clinical Reasoning: Past History and Current Trends'. *Medical Education* 39 (4), 418–427

Norman, G., Sherbino, J., Dore, K., Wood, T., Young, M., Gaissmaier, W., Kreuger, S.,

and Monteiro, S. (2014) 'The Etiology of Diagnostic Errors'. *Academic Medicine* 89 (2), 277–284

Norman, G.R., Monteiro, S.D., Sherbino, J., Ilgen, J.S., Schmidt, H.G., and Mamede, S. (2017) 'The Causes of Errors in Clinical Reasoning: Cognitive Biases, Knowledge Deficits, and Dual Process Thinking'. in *Academic Medicine*. vol. 92 (1). 23–30

O'cathain, A., Murphy, E., and Nicholl, J. (2008) 'The Quality of Mixed Methods Studies in Health Services Research'. *Journal of Health Services Research & Policy* 13 (2), 92–98

Offredy, M. and Vickers, P.S. (2010) *Developing a Healthcare Research Proposal : An Interactive Student Guide*. Wiley-Blackwell

Oliver, A., Tortajada, M., Lladó, X., Freixenet, J., Ganau, S., Tortajada, L., Vilagran, M., Sentís, M., and Martí, R. (2015) 'Breast Density Analysis Using an Automatic Density Segmentation Algorithm'. *Journal of Digital Imaging* 28 (5), 604–612

Oliver, P. (2010) *Student's Guide to Research Ethics*. 2nd edn. Open University Press

Onega, T., Beaber, E.F., Sprague, B.L., Barlow, W.E., Haas, J.S., Tosteson, A.N.A., D. Schnall, M., Armstrong, K., Schapira, M.M., Geller, B., Weaver, D.L., and Conant, E.F. (2014) 'Breast Cancer Screening in an Era of Personalized Regimens: A Conceptual Model and National Cancer Institute Initiative for Risk-Based and Preference-Based Approaches at a Population Level'. *Cancer* 120 (19), 2955–2964

Ostroff, C., Kinicki, A.J., and Muhammad, R.S. (2013) 'Organizational Culture and Climate'. in *Handbook of Psychology*. 2nd edn. Wiley, 643–676

Otten, J.D.M., Karssemeijer, N., Hendriks, J.H.C.L., Groenewoud, J.H., Fracheboud, J., Verbeek, A.L.M., de Koning, H.J., and Holland, R. (2005) 'Effect of Recall Rate on Earlier Screen Detection of Breast Cancers Based on the Dutch Performance Indicators'. *JNCI: Journal of the National Cancer Institute* 97 (10), 748–754

Owen, G. (1989) 'Collective Decision-Making: An Economic Outlook : Shmuel Nitzan and Jacob Paroush, Cambridge University Press, Cambridge, 1986) Pp. X+132, $34.50'. *Journal of Economic Behavior & Organization* 11 (2), 302–303

Parahoo, K. (2014) *Nursing Research : Principles, Process and Issues*. 3rd edn. Palgrave

Parekh, V.S. and Jacobs, M.A. (2017) 'Integrated Radiomic Framework for Breast Cancer and Tumor Biology Using Advanced Machine Learning and Multiparametric MRI'. *Npj Breast Cancer* 3 (1), 43

Patel, V.L., Kaufman, D.R., and Arocha, J.F. (2002) 'Emerging Paradigms of Cognition in Medical Decision-Making.' *Journal of Biomedical Informatics* 35 (1), 52–75

Patton, M.Q. (2000) *Qualitative Research & Evaluation Methods : Integrating Theory and Practice*. Fourth. SAGE Publications, Inc.

Pauli, R., Hammond, S., Cooke, J., and Ansell, J. (1996) 'Comparison of

Radiographer/Radiologist Double Film Reading with Single Reading in Breast Cancer Screening'. *Journal of Medical Screening* 3 (1), 18–22

Pawson, R. (2013) *The Science of Evaluation: A Realist Manifesto* -. London: Sage Publications

Pearson, H. (2013) 'Science and Intuition: Do Both Have a Place in Clinical Decision Making?' *British Journal of Nursing* 22 (4), 212–215

Pennycook, G., Fugelsang, J.A., and Koehler, D.J. (2015) 'What Makes Us Think? A Three-Stage Dual-Process Model of Analytic Engagement'. *Cognitive Psychology* 80, 34–72

Pernet, C.R., Wilcox, R., and Rousselet, G.A. (2012) 'Robust Correlation Analyses: False Positive and Power Validation Using a New Open Source Matlab Toolbox.' *Frontiers in Psychology* 3, 606

Perry, N., Broeders, M., de Wolf, C., Tornberg, S., Holland, R., and von Karsa, L. (2007) 'European Guidelines for Quality Assurance in Breast Cancer Screening and Diagnosis. Fourth Edition--Summary Document'. *Annals of Oncology* 19 (4), 614–622

Phillips, M., Marsden, H., Jaffe, W., Matin, R.N., Wali, G.N., Greenhalgh, J., McGrath, E., James, R., Ladoyanni, E., Bewley, A., Argenziano, G., and Palamaras, I. (2019) 'Assessment of Accuracy of an Artificial Intelligence Algorithm to Detect Melanoma in Images of Skin Lesions'. *JAMA Network Open* 2 (10), e1913436

Pieterse, T., Lawrence, H., and Friedrich-Nel, H. (2016) 'Critical Thinking Ability of 3rd Year Radiography Students'. *Health SA Gesondheid* 21, 381–390

Pisano, E.D. (2020) 'AI System for Breast Cancer Screening'. *Nature Magazine* 35–36

Pisano, E.D., Hendrick, R.E., Yaffe, M.J., Baum, J.K., Acharyya, S., Cormack, J.B., Hanna, L.A., Conant, E.F., Fajardo, L.L., Bassett, L.W., D'Orsi, C.J., Jong, R.A., Rebner, M., Tosteson, A.N.A., Gatsonis, C.A., and DMIST Investigators Group (2008) 'Diagnostic Accuracy of Digital versus Film Mammography: Exploratory Analysis of Selected Population Subgroups in DMIST'. *Radiology* 246 (2), 376–383

Pluye, P. and Hong, Q.N. (2014) 'Combining the Power of Stories and the Power of Numbers: Mixed Methods Research and Mixed Studies Reviews'. *Annual Review of Public Health* 35 (1), 29–45

Polit, D.F. and Beck, C.T. (2016) *Nursing Research : Generating and Assessing Evidence for Nursing Practice*. 10th edn. Lippincott Williams and Wilkins

Polit, D.F. and Tatano Beck, C. (2010) *Essentials of Nursing Research Appraising Evidence for Nursing Practice*. 7th edn. Lippincott Williams & Wilkins

Posso, M., Carles, M., Rué, M., Puig, T., and Bonfill, X. (2016) 'Cost-Effectiveness of Double Reading versus Single Reading of Mammograms in a Breast Cancer Screening Programme'. *PLOS ONE* 11 (7), e0159806

Pow, R.E., Mello-Thoms, C., and Brennan, P. (2016) 'Evaluation of the Effect of Double Reporting on Test Accuracy in Screening and Diagnostic Imaging Studies: A Review of the Evidence'. *Journal of Medical Imaging and Radiation Oncology* 60 (3), 306–314

Price, R.C. and Edwards, H.M. (2008) 'Harnessing Competence and Confidence: Dimensions in Education and Development for Advanced and Consultant Practice'. *Radiography* 14 (SUPPL. 1), 65–70

Public Health England (2017a) *NHS Breast Screening Programme Consolidated Standards Public Health England Leads the NHS Screening Programmes*.

Public Health England (2017b) *NHS Breast Screening Programme Reporting, Classification and Monitoring of Interval Cancers and Cancers Following Previous Assessment*. (August), 1–20

Public Health England (2016) *NHS Breast Screening Programme Guidance on Who Can Undertake Arbitration Public Health England Leads the NHS Screening Programmes*. (August), 1–6

Public Health England and PHE (2011) *Quality Assurance Guidelines for Breast Cancer Screening Radiology (NHSBSP Publication No 59)*.

Quach, K. (2018) *IBM Watson Dishes out 'Dodgy Cancer Advice', Google Translate Isn't Better than Humans yet, and Other AI Tidbits • The Register*.

Rajan, S. and Sharma, N. (2016) 'British Society of Breast Radiology Annual Scientific Meeting 2016'. *Breast Cancer Research* 18 (S1), 107

Rattray, J. and Jones, M.C. (2007) 'Essential Elements of Questionnaire Design and Development.' *Journal of Clinical Nursing* 16 (2), 234–243

Rebar, C.R. and Gersch, C.J. (2014) *Understanding Nursing Research : Using Research in Evidence-Based Practice*. 4th edn. Lippincott Williams & Wilkins

Redelmeier, D.A., Ferris, L.E., Tu, J. V, Hux, J.E., and Schull, M.J. (2001) 'Problems for Clinical Judgement: Introducing Cognitive Psychology as One More Basic Science.' *CMAJ : Canadian Medical Association Journal = Journal de l'Association Medicale Canadienne* 164 (3), 358–60

Richards, H.M., Schwartz, L.J., Hm, R., and Ethics, S.L. (2002) *Ethics of Qualitative Research: Are There Special Issues for Health Services Research?* vol. 19

Ritchie, Jane. and Lewis, J. (2003) *Qualitative Research Practice : A Guide for Social Science Students and Researchers*. Sage Publications

Ritchie, Jane and Lewis, J. (2003) *QUALITATIVE RESEARCH PRACTICE A Guide for Social Science Students and Researchers*. 1st edn. SAGE Publications Ltd

Ritchie, J., Lewis, J., McNaughton Nicholls, C., and Ormston, R. (2014) *Qualitative Research Practice : A Guide for Social Science Students and Researchers*. Second edi.

Roberts, P., Priest, H., and Traynor, M. (2006) 'Reliability and Validity in Research'. *Nursing Standard* 20 (44), 41–45

Rodríguez-Ruiz, A., Krupinski, E., Mordang, J.-J., Schilling, K., Heywang-Köbrunner, S.H., Sechopoulos, I., Mann, R.M., and Rodríguez-Ruiz, R; Krupinski, A; Mordang, E; Schilling, J; Heywang-Köbrunner' K; Sechopoulos IMann, S. (2019) 'Detection of Breast Cancer with Mammography: Effect of an Artificial Intelligence Support System'. *Radiology* 290 (2), 305–314

Rodriguez-Ruiz, A., Lång, K., Gubern-Merida, A., Broeders, M., Gennaro, G., Clauser, P., Helbich, T.H., Chevalier, M., Tan, T., Mertelmeier, T., Wallis, M.G., Andersson, I., Zackrisson, S., Mann, R.M., and Sechopoulos, I. (2019) 'Stand-Alone Artificial Intelligence for Breast Cancer Detection in Mammography: Comparison With 101 Radiologists'. *Journal of the National Cancer Institute* 111 (9), 916–922

Rolstad, S., Adler, J., and Rydén, A. (2011) 'Response Burden and Questionnaire Length: Is Shorter Better? A Review and Meta-Analysis'. *Value in Health* 14 (8), 1101–1108

Roman, M., Hubbard, R.A., Sebuodegard, S., Miglioretti, D.L., Castells, X., and Hofvind, S. (2013) 'The Cumulative Risk of False-Positive Results in the Norwegian Breast Cancer Screening Program: Updated Results'. *Cancer* 119 (22), 3952–3958

Rose Jemutai, C. and Wambua, P.P. (2016) 'Integrated Marketing Communication and Performance of Kenya Post and Savings Bank'. *Social Sciences* 5 (3), 37

Rose, S.L., Tidwell, A.L., Bujnoch, L.J., Kushwaha, A.C., Nordmann, A.S., and Sexton, R. (2013) 'Implementation of Breast Tomosynthesis in a Routine Screening Practice: An Observational Study'. *American Journal of Roentgenology* 200 (6), 1401–1408

Roth, L.M. and Markova, T. (2012) 'Essentials for Great Teams: Trust, Diversity, Communication ... and Joy.' *Journal of the American Board of Family Medicine : JABFM* 25 (2), 146–8

Rothstein A, B.C.T. (2004) *On Conflict and Consensus: A Handbook on Formal Consensus Decision Making*. 2nd edn. Gavin White

Russell, C.J. and Bobko, P. (1992) 'Moderated Regression Analysis and Likert Scales: Too Coarse for Comfort.' *Journal of Applied Psychology* 77 (3), 336–342

Rutter, C.M. and Taplin, S. (2000) 'Assessing Mammographers' Accuracy: A Comparison of Clinical and Test Performance'. *Journal of Clinical Epidemiology* 53 (5), 443–450

Salant, P. and Dillman, D.A. (1994) *How to Conduct Your Own Survey*. Wiley

Saldaña, J. (2015) *The Coding Manual for Qualitative Researchers*. 3rd edn. SAGE Publications Ltd

Salim, M., Wåhlin, E., Dembrower, K., Azavedo, E., Foukakis, T., Liu, Y., Smith, K., Eklund, M., and Strand, F. (2020) 'External Evaluation of 3 Commercial Artificial

Intelligence Algorithms for Independent Assessment of Screening Mammograms'. *JAMA Oncology* 1–8

Sandhu, R., Parker, J.S., Jones, W.D., Livasy, C.A., and Coleman, W.B. (2010) 'Microarray-Based Gene Expression Profiling for Molecular Classification of Breast Cancer and Identification of New Targets for Therapy'. *Laboratory Medicine* 41 (6), 364–372

Sanjari, M., Bahramnezhad, F., Fomani, F.K., Shoghi, M., and Cheraghi, M.A. (2014) 'Ethical Challenges of Researchers in Qualitative Studies: The Necessity to Develop a Specific Guideline.' *Journal of Medical Ethics and History of Medicine* 7, 14

Sankatsing, V.D. V., Juraniec, K., Grimm, S.E., Joore, M.A., Pijnappel, R.M., de Koning, H.J., and van Ravesteyn, N.T. (2020) 'Cost-Effectiveness of Digital Breast Tomosynthesis in Population-Based Breast Cancer Screening: A Probabilistic Sensitivity Analysis'. *Radiology* 192505

Santiago-Delefosse, M., Gavin, A., Bruchez, C., Roux, P., and Stephen, S.L. (2016) 'Quality of Qualitative Research in the Health Sciences: Analysis of the Common Criteria Present in 58 Assessment Guidelines by Expert Users'. *Social Science & Medicine* 148, 142–151

Saunders, M.N.K., Lewis, P., and Thornhill, A. (2009) *Research Methods for Business Students*. Prentice Hall

Schaffter, T., Buist, D.S.M., Lee, C.I., Nikulin, Y., Ribli, D., Guan, Y., Lotter, W., Jie, Z., Du, H., Wang, S., Feng, J., Feng, M., Kim, H.E., Albiol, F., Albiol, A., Morrell, S., Wojna, Z., Ahsen, M.E., Asif, U., Jimeno Yepes, A., Yohanandan, S., Rabinovici-Cohen, S., Yi, D., Hoff, B., Yu, T., Chaibub Neto, E., Rubin, D.L., Lindholm, P., Margolies, L.R., McBride, R.B., Rothstein, J.H., Sieh, W., Ben-Ari, R., Harrer, S., Trister, A., Friend, S., Norman, T., Sahiner, B., Strand, F., Guinney, J., Stolovitzky, G., Mackey, L., Cahoon, J., Shen, L., Sohn, J.H., Trivedi, H., Shen, Y., Buturovic, L., Pereira, J.C., Cardoso, J.S., Castro, E., Kalleberg, K.T., Pelka, O., Nedjar, I., Geras, K.J., Nensa, F., Goan, E., Koitka, S., Caballero, L., Cox, D.D., Krishnaswamy, P., Pandey, G., Friedrich, C.M., Perrin, D., Fookes, C., Shi, B., Cardoso Negrie, G., Kawczynski, M., Cho, K., Khoo, C.S., Lo, J.Y., Sorensen, A.G., and Jung, H. (2020) 'Evaluation of Combined Artificial Intelligence and Radiologist Assessment to Interpret Screening Mammograms'. *JAMA Network Open* 3 (3), e200265

Schein, E.H. (2010) *Organizational Culture and Leadership*. Jossey-Bass

Schell, M.J., Yankaskas, B.C., Ballard-Barbash, R., Qaqish, B.F., Barlow, W.E., Rosenberg, R.D., and Smith-Bindman, R. (2007) 'Evidence-Based Target Recall Rates for Screening Mammography 1'. *Radiology* 243 (3), 681–689

Schermerhorn, J.R. (2012) *Organizational Behavior.* John Wiley & Sons; International student edition

Schiff, G.D., Hasan, O., Kim, S., Abrams, R., Cosby, K., Lambert, B.L., Elstein, A.S., Hasler, S., Kabongo, M.L., Krosnjar, N., Odwazny, R., Wisniewski, M.F., and

McNutt, R.A. (2009) 'Diagnostic Error in Medicine'. *Archives of Internal Medicine* 169 (20), 1881–1887

Schmidt, H.G. and Mamede, S. (2015) 'How to Improve the Teaching of Clinical Reasoning: A Narrative Review and a Proposal'. *Medical Education* 49 (10), 961–973

Schmidt, H.G., Mamede, S., van den Berge, K., van Gog, T., van Saase, J.L.C.M., and Rikers, R.M.J.P. (2014) 'Exposure to Media Information About a Disease Can Cause Doctors to Misdiagnose Similar-Looking Clinical Cases'. *Academic Medicine* 89 (2), 285–291

Schmidt, H.G. and Rikers, R.M.J.P. (2007) 'How Expertise Develops in Medicine: Knowledge Encapsulation and Illness Script Formation'. *Medical Education* 41 (12), 1133–1139

Schneider, A., Hommel, G., and Blettner, M. (2010) 'Linear Regression Analysis: Part 14 of a Series on Evaluation of Scientific Publications.' *Deutsches Arzteblatt International* 107 (44), 776–82

Schneider, B., Ehrhart, M.G., and Macey, W.H. (2013) 'Organizational Climate and Culture'. *Annual Review of Psychology* 64 (1), 361–388

Schousboe, J.T., Kerlikowske, K., Loh, A., and Cummings, S.R. (2011) 'Personalizing Mammography by Breast Density and Other Risk Factors for Breast Cancer: Analysis of Health Benefits and Cost-Effectiveness'. *Annals of Internal Medicine*

155 (1), 10–20

Schroder, C., Medves, J., Paterson, M., Byrnes, V., Chapman, C., O'Riordan, A., Pichora, D., and Kelly, C. (2011) 'Development and Pilot Testing of the Collaborative Practice Assessment Tool'. *Journal of Interprofessional Care* 25 (3), 189–195

Scott, H.J., Evans, A., Gale, A.G., Murphy, A., and Reed, J. (2009) 'The Relationship between Real Life Breast Screening and an Annual Self Assessment Scheme'. in *Proc.SPIE*. held 12 March 2009

Scott, H.J. and Gale, A.G. (2007) *How Much Is Enough? Factors Affecting the Optimal Interpretation of Breast Screening Mammograms*. 360

Scragg, B., Shaikh, S., Shires, G., Stein Hodgins, J., Mercer, C., Robinson, L., and Wray, J. (2017) 'An Exploration of Mammographers' Attitudes towards the Use of Social Media for Providing Breast Screening Information to Clients'. *Radiography* 23 (3), 249–255

Searle, R. and Skinner, D. (2011) 'Trust and Human Resource Management'. in *Trust and Human Resource Management*. Edward Elgar, 42–64

Searle, S.R. (Shayle R.. (2006) *Linear Models for Unbalanced Data*. Wiley-Interscience

Sechopoulos, I., Teuwen, J., and Mann, R. (2020) 'Artificial Intelligence for Breast Cancer Detection in Mammography and Digital Breast Tomosynthesis: State of

the Art'. *Seminars in Cancer Biology* (November 2019)

Seely, J.M. and Alhassan, T. (2018) 'Screening for Breast Cancer in 2018—What Should We Be Doing Today?' *Current Oncology* 25 (June), S115–S124

Senior, B. and Swailes, S. (2016) *Organizational Change*. 5th edn. FT Publishing International

Shapley, L. and Grofman, B. (1984) *Optimizing Group Judgmental Accuracy in the Presence of Interdependencies**. vol. 43

Sharma, N. (2018) 'Special Issue on Breast Imaging: Part 1.' *Clinical Radiology* 73 (4), 325–326

Sherbino, J., Dore, K.L., Wood, T.J., Young, M.E., Gaissmaier, W., Kreuger, S., and Norman, G.R. (2012) 'The Relationship Between Response Time and Diagnostic Accuracy'. *Academic Medicine* 87 (6), 785–791

Shumate, M., Ibrahim, R., and Levitt, R. (2010) 'Dynamic Information Retrieval and Allocation Flows in Project Teams with Discontinuous Membership'. *European J. of International Management* 4 (6), 556

Sim, J, Wright, C. (2000) *Research in Health Care: Concepts, Designs and Methods.* vol. 6. Stanley Thornes Ltd

Simons, T.L. and Peterson, R.S. (2000) 'Task Conflict and Relationship Conflict in Top Management Teams: The Pivotal Role of Intragroup Trust'. *Journal of Applied*

Psychology 85 (1), 102–111

Singh, H., Schiff, G.D., Graber, M.L., Onakpoya, I., and Thompson, M.J. (2017) 'The Global Burden of Diagnostic Errors in Primary Care.' *BMJ Quality & Safety* 26 (6), 484–494

Sinn, H.P. and Kreipe, H. (2013) 'A Brief Overview of the WHO Classification of Breast Tumors, 4th Edition, Focusing on Issues and Updates from the 3rd Edition'. *Breast Care* 8 (2), 149–154

Skaane, P., Bandos, A.I., Gullien, R., Eben, E.B., Ekseth, U., Haakenaasen, U., Izadi, M., Jebsen, I.N., Jahr, G., Krager, M., and Hofvind, S. (2013a) 'Prospective Trial Comparing Full-Field Digital Mammography (FFDM) versus Combined FFDM and Tomosynthesis in a Population-Based Screening Programme Using Independent Double Reading with Arbitration'. *European Radiology* 23 (8), 2061–2071

Skaane, P., Bandos, A.I., Gullien, R., Eben, E.B., Ekseth, U., Haakenaasen, U., Izadi, M., Jebsen, I.N., Jahr, G., Krager, M., Niklason, L.T., Hofvind, S., and Gur, D. (2013b) 'Comparison of Digital Mammography Alone and Digital Mammography Plus Tomosynthesis in a Population-Based Screening Program'. *Radiology* 267 (1), 47–56

Skaane, P., Diekmann, F., Balleyguier, C., Diekmann, S., Piguet, J.-C., Young, K., Abdelnoor, M., and Niklason, L. (2008) 'Observer Variability in Screen-Film Mammography versus Full-Field Digital Mammography with Soft-Copy Reading'. *European Radiology* 18 (6), 1134–1143

Skaane, P., Sebuødegård, S., Bandos, A.I., Gur, D., Østerås, B.H., Gullien, R., and Hofvind, S. (2018) 'Performance of Breast Cancer Screening Using Digital Breast Tomosynthesis: Results from the Prospective Population-Based Oslo Tomosynthesis Screening Trial'. *Breast Cancer Research and Treatment* 169 (3), 489–496

Smith-Bindman, R., Chu, P.W., Miglioretti, D.L., Sickles, E.A., Blanks, R., Ballard-Barbash, R., Bobo, J.K., Lee, N.C., Wallis, M.G., Patnick, J., and Kerlikowske, K. (2003) 'Comparison of Screening Mammography in the United States and the United Kingdom'. *JAMA* 290 (16), 2129

Soh, B.P., Lee, W.B., Mello-Thoms, C., Tapia, K., Ryan, J., Hung, W.T., Thompson, G., Heard, R., and Brennan, P. (2015) 'Certain Performance Values Arising from Mammographic Test Set Readings Correlate Well with Clinical Audit'. *Journal of Medical Imaging and Radiation Oncology* 59 (4), 403–410

Sokolovskaya, E., Shinde, T., Ruchman, R.B., Kwak, A.J., Lu, S., Shariff, Y.K., Wiggins, E.F., and Talangbayan, L. (2015) 'The Effect of Faster Reporting Speed for Imaging Studies on the Number of Misses and Interpretation Errors: A Pilot Study.' *Journal of the American College of Radiology : JACR* 12 (7), 683–688

Soll, J.B. (1985) *Organizational Behavior and Human Decision Processes.* vol. 65. Academic Press

Song, H., Chien, A.T., Fisher, J., Martin, J., Peters, A.S., Hacker, K., Rosenthal, M.B., and Singer, S.J. (2015) 'Development and Validation of the Primary Care Team

Dynamics Survey'. *Health Services Research* 50 (3), 897–921

Sorkin, R.D., Hays, C.J., West, R., and Sorkin, D. (2001) *Signal-Detection Analysis of Group Decision Making*. vol. 108

Sprague, B.L., Coley, R.Y., Kerlikowske, K., Rauscher, G.H., Henderson, L.M., Onega, T., Lee, C.I., Herschorn, S.D., Tosteson, A.N.A., and Miglioretti, D.L. (2020) 'Assessment of Radiologist Performance in Breast Cancer Screening Using Digital Breast Tomosynthesis vs Digital Mammography.' *JAMA Network Open* 3 (3), e201759

Stanek, P. (2013) 'Theoretical Aspects of Collective Decision Making - Survey of the Economic Literature'. *Comparative Economic Research* 16 (1), 103–121

Stanovich, K.E. (2009) *What Intelligence Tests Miss : The Psychology of Rational Thought*. Yale University Press

Stark, M. and Fins, J.J. (2014) 'The Ethical Imperative to Think about Thinking'. *Cambridge Quarterly of Healthcare Ethics* 23 (04), 386–396

Stec, N., Arje, D., Moody, A.R., Krupinski, E.A., and Tyrrell, P.N. (2018) 'A Systematic Review of Fatigue in Radiology: Is It a Problem?' *American Journal of Roentgenology* 210 (4), 799–806

Steel, J. (n.d.) 'Third-Reading in the NHS Breast Screening Programme. How Arbitrary Is Arbitration?' in *Symposium Mammographicum Conference 3-5 July ACC*

Liverpool.

Stephens, N. (2007) 'Collecting Data from Elites and Ultra Elites: Telephone and Face-to-Face Interviews with Macroeconomists'. *Qualitative Research* 7 (2), 203–216

Stewart, B.W., Wild, C., International Agency for Research on Cancer, and World Health Organization (n.d.) *World Cancer Report 2014*.

Stiegler, M.P. and Ruskin, K.J. (2012) 'Decision-Making and Safety in Anesthesiology'. *Current Opinion in Anaesthesiology* 25 (6), 724–729

Stockman, C. (2015) 'Electronic Journal of Business Research Methods : EJBRM.' *ECRM.* 13 (2), 74–84

Subar, A.F., Ziegler, R.G., Thompson, F.E., Johnson, C.C., Weissfeld, J.L., Reding, D., Kavounis, K.H., Hayes, R.B., and Prostate, Lung, Colorectal, and O.C.S.T.I. (2001) 'Is Shorter Always Better? Relative Importance of Questionnaire Length and Cognitive Ease on Response Rates and Data Quality for Two Dietary Questionnaires.' *American Journal of Epidemiology* 153 (4), 404–9

Suter, M.B., Pesapane, F., Agazzi, G.M., Gagliardi, T., Nigro, O., Bozzini, A., Priolo, F., Penco, S., Cassano, E., Chini, C., and Squizzato, A. (2020) 'Diagnostic Accuracy of Contrast-Enhanced Spectral Mammography for Breast Lesions: A Systematic Review and Meta-Analysis'. *Breast* 53, 8–17

Svedbo Engström, M., Leksell, J., Johansson, U.-B., Eeg-Olofsson, K., Borg, S., Palaszewski, B., and Gudbjörnsdottir, S. (2018) 'A Disease-Specific Questionnaire for Measuring Patient-Reported Outcomes and Experiences in the Swedish National Diabetes Register: Development and Evaluation of Content Validity, Face Validity, and Test-Retest Reliability'. *Patient Education and Counseling* 101 (1), 139–146

Tabár, L., Fagerberg, C.J., Gad, A., Baldetorp, L., Holmberg, L.H., Gröntoft, O., Ljungquist, U., Lundström, B., Månson, J.C., and Eklund, G. (1985) 'Reduction in Mortality from Breast Cancer after Mass Screening with Mammography. Randomised Trial from the Breast Cancer Screening Working Group of the Swedish National Board of Health and Welfare.' *Lancet (London, England)* 1 (8433), 829–32

Tabàr, L., Fagerberg, G., Duffy, S.W., Day, N.E., Gad, A., and Gröntoft, O. (1992) 'Update of the Swedish Two-County Program of Mammographic Screening for Breast Cancer.' *Radiologic Clinics of North America* 30 (1), 187–210

Tagliafico, A.S., Calabrese, M., Bignotti, B., Signori, A., Fisci, E., Rossi, F., Valdora, F., and Houssami, N. (2017) 'Accuracy and Reading Time for Six Strategies Using Digital Breast Tomosynthesis in Women with Mammographically Negative Dense Breasts'. *European Radiology* 27 (12), 5179–5184

Taplin, S.H., Weaver, S., Chollette, V., Marks, L.B., Jacobs, A., Schiff, G., Stricker, C.T., Bruinooge, S.S., and Salas, E. (2015) 'Teams and Teamwork during a Cancer

Diagnosis: Interdependency within and between Teams.' *Journal of Oncology Practice* 11 (3), 231–8

Tashakkori, A. and Teddlie, C. (2010) *Sage Handbook of Mixed Methods in Social & Behavioral Research*. 2nd edn. SAGE Publications

Taylor-Phillips, S., Jenkinson, D., Stinton, C., Wallis, M.G., Dunn, J., and Clarke, A. (2018) 'Double Reading in Breast Cancer Screening: Cohort Evaluation in the CO-OPS Trial'. *Radiology* 287 (3), 749–757

Taylor-Phillips, S., Wallis, M.G., Jenkinson, D., Adekanmbi, V., Parsons, H., Dunn, J., Stallard, N., Szczepura, A., Gates, S., Kearins, O., Duncan, A., Hudson, S., and Clarke, A. (2016) 'Effect of Using the Same vs Different Order for Second Readings of Screening Mammograms on Rates of Breast Cancer Detection'. *JAMA* 315 (18), 1956–1965

Taylor, E., Hewitt, K., Dr, R.A.R., Dr, S.H.H., and Lawless, W.F. (2013) 'Group Decision-Making: Consensus Rule Versus Majority Rule'. *Procedia Technology* 9, 498–504

Teddlie, C. and Tashakkori, A. (2009) *Foundations of Mixed Methods Research : Integrating Quantitative and Qualitative Approaches in the Social and Behavioral Sciences*. SAGE

Thammasitboon, S. and Cutrer, W.B. (2013) 'Diagnostic Decision-Making and Strategies to Improve Diagnosis'. *Current Problems in Pediatric and Adolescent Health Care* 43 (9), 232–241

The Royal College of Radiologists (2020) 'Clinical Radiology UK Workforce Census 2019 Report'. *Clinical Radiology* (April)

Théberge, I., Chang, S.-L., Vandal, N., Daigle, J.-M., Guertin, M.-H., Pelletier, É., and Brisson, J. (2014) 'Radiologist Interpretive Volume and Breast Cancer Screening Accuracy in a Canadian Organized Screening Program'. *JNCI: Journal of the National Cancer Institute* 106 (3)

Thibault, G.E. (2013) 'Reforming Health Professions Education Will Require Culture Change and Closer Ties between Classroom and Practice'. *Health Affairs* 32 (11), 1928–1932

Thomas, E.J., Sexton, J.B., and Helmreich, R.L. (2003) 'Discrepant Attitudes about Teamwork among Critical Care Nurses and Physicians*'. *Critical Care Medicine* 31 (3), 956–959

Thompson, C. and Dowding, D. (2002) *Clinical Decision Making and Judgement in Nursing*. Churchill Livingstone

Tice, J.A., Cummings, S.R., Ziv, E., and Kerlikowske, K. (2005) 'Mammographic Breast Density and the Gail Model for Breast Cancer Risk Prediction in a Screening Population'. *Breast Cancer Research and Treatment* 94 (2), 115–122

Todres, L. and Galvin, K. (2012) '"In The Middle Of Everywhere:" A Phenomenological Study Of Mobility And Dwelling Amongst Rural Elders'. *Phenomenology and Practice* 6 (1), 55–68

Topol, E. (2019) *Preparing the Healthcare Workforce to Deliver the Digital Future The Topol Review An Independent Report on Behalf of the Secretary of State for Health and Social Care*.

Torres-Mejía, G., Smith, R.A., Carranza-Flores, M. de la L., Bogart, A., Martínez-Matsushita, L., Miglioretti, D.L., Kerlikowske, K., Ortega-Olvera, C., Montemayor-Varela, E., Angeles-Llerenas, A., Bautista-Arredondo, S., Sánchez-González, G., Martínez-Montañez, O.G., Uscanga-Sánchez, S.R., Lazcano-Ponce, E., and Hernández-Ávila, M. (2015) 'Radiographers Supporting Radiologists in the Interpretation of Screening Mammography: A Viable Strategy to Meet the Shortage in the Number of Radiologists'. *BMC Cancer* 15 (1), 410

Trimble, M and Hamilton, P. (2016) 'The Thinking Doctor: Clinical Decision Making in Contemporary Medicine'. *Clinical Medicine* 16 (4), 343–349

Trimble, Michael and Hamilton, P. (2016) 'The Thinking Doctor: Clinical Decision Making in Contemporary Medicine'. *Clinical Medicine, Journal of the Royal College of Physicians of London* 16 (4), 343–346

Trippas, D., Handley, S.J., Verde, M.F., and Morsanyi, K. (2016) 'Logic Brightens My Day: Evidence for Implicit Sensitivity to Logical Validity.' *Journal of Experimental Psychology: Learning, Memory, and Cognition* 42 (9), 1448–1457

Trister, A.D., Buist, D.S.M., and Lee, C.I. (2017) 'Will Machine Learning Tip the Balance in Breast Cancer Screening?' *JAMA Oncology* 3 (11), 1463

Trotter, R.T. (2012) 'Qualitative Research Sample Design and Sample Size: Resolving and Unresolved Issues and Inferential Imperatives'. *Preventive Medicine* 55 (5), 398–400

Tsalatsanis, A., Hozo, I., Kumar, A., and Djulbegovic, B. (2015) 'Dual Processing Model for Medical Decision-Making: An Extension to Diagnostic Testing'. *PLoS ONE* 10 (8), 1–17

Turnbull, C. and Rahman, N. (2008) 'Genetic Predisposition to Breast Cancer: Past, Present, and Future'. *Annual Review of Genomics and Human Genetics* 9 (1), 321–345

Újhelyi, M., Pukancsik, D., Kelemen, P., Kovács, E., Kenessey, I., Udvarhelyi, N., Bak, M., Kovács, T., and Mátrai, Z. (2016) 'Does Breast Screening Offer a Survival Benefit? A Retrospective Comparative Study of Oncological Outcomes of Screen-Detected and Symptomatic Early Stage Breast Cancer Cases'. *European Journal of Surgical Oncology (EJSO)* 42 (12), 1814–1820

UK Policy Framework for Health and Social Care Research - Health Research Authority (n.d.)

Valentine, M.A., Nembhard, I.M., and Edmondson, A.C. (2015) 'Measuring Teamwork in Health Care Settings'. *Medical Care* 53 (4), 16–30

Vashdi, D.R., Bamberger, P.A., and Erez, M. (2013) 'Can Surgical Teams Ever Learn? The Role of Coordination, Complexity, and Transitivity in Action Team Learning'.

Academy of Management Journal 56 (4), 945–971

Vatanpour, H., Khorramnia, A., and Forutan, N. (2013) 'Silo Effect a Prominence Factor to Decrease Efficiency of Pharmaceutical Industry'. *Iranian Journal of Pharmaceutical Research* 12 (SUPPL.), 203–212

Velde, B.P., Wittman, P.P., and Vos, P. (2006) 'Development of Critical Thinking in Occupational Therapy Students.' *Occupational Therapy International* 13 (1), 49–60

van de Vijver, M.J., He, Y.D., van 't Veer, L.J., Dai, H., Hart, A.A.M., Voskuil, D.W., Schreiber, G.J., Peterse, J.L., Roberts, C., Marton, M.J., Parrish, M., Atsma, D., Witteveen, A., Glas, A., Delahaye, L., van der Velde, T., Bartelink, H., Rodenhuis, S., Rutgers, E.T., Friend, S.H., and Bernards, R. (2002) 'A Gene-Expression Signature as a Predictor of Survival in Breast Cancer'. *New England Journal of Medicine* 347 (25), 1999–2009

Vila-Henninger, L.A. (2015) 'Toward Defining the Causal Role of Consciousness: Using Models of Memory and Moral Judgment from Cognitive Neuroscience to Expand the Sociological Dual-Process Model'. *Journal for the Theory of Social Behaviour* 45 (2), 238–260

Vinnicombe, S.J. (2018) 'Breast Density: Why All the Fuss?' *Clinical Radiology* 73 (4), 334–357

Wadhwa, A., Sullivan, J.R., and Gonyo, M.B. (2016) 'Missed Breast Cancer: What Can

We Learn?' *Current Problems in Diagnostic Radiology* 45 (6), 402–419

Waite, S., Scott, J., Gale, B., Fuchs, T., Kolla, S., and Reede, D. (2017) 'Interpretive Error in Radiology.' *AJR. American Journal of Roentgenology* 208 (4), 739–749

Walters, S.J., Stern, C., and Robertson-Malt, S. (2016) 'The Measurement of Collaboration within Healthcare Settings: A Systematic Review of Measurement Properties of Instruments.' *JBI Database of Systematic Reviews and Implementation Reports* 14 (4), 138–97

Watanabe, AT Mayo, RC, Chang Sen, LQ, Kapoor, M, Leung, J. (2019) *Artificial Intelligence Software to Improve Mammography Workflow*.

Weaver, S.J., Dy, S.M., and Rosen, M.A. (2014) 'Team-Training in Healthcare: A Narrative Synthesis of the Literature'. *BMJ Quality & Safety* 23 (5), 359–372

Weber, E.U., Böckenholt, U., Hilton, D.J., and Wallace, B. (1993) 'Determinants of Diagnostic Hypothesis Generation: Effects of Information, Base Rates, and Experience.' *Journal of Experimental Psychology. Learning, Memory, and Cognition* 19 (5), 1151–64

Weedon-Fekjær, H., Romundstad, P.R., and Vatten, L.J. (2014) 'Modern Mammography Screening and Breast Cancer Mortality: Population Study.' *BMJ (Clinical Research Ed.)* 348, g3701

Weisberg, S. (2014) *Applied Linear Regression*. ed. by Hoboken, N. John Wiley &

Sons; Inc

Weiss, S.J. and Davis, H.P. (1985) 'Validity and Reliability of the Collaborative Practice Scales.' *Nursing Research* 34 (5), 299–305

Welch, H.G. and Passow, H.J. (2014) 'Quantifying the Benefits and Harms of Screening Mammography'. *JAMA Internal Medicine* 174 (3), 448–454

Wells, J.. and Cooke, J. (1996) 'Film Reading Practice of UK Breast Screening Units'. *The Breast* 5 (6), 404–409

Wendt, C. and Margolin, S. (2019) 'Identifying Breast Cancer Susceptibility Genes–a Review of the Genetic Background in Familial Breast Cancer'. *Acta Oncologica* 58 (2), 135–146

West, M.A. and Lyubovnikova, J. (2012) 'Real Teams or Pseudo Teams? The Changing Landscape Needs a Better Map'. *Industrial and Organizational Psychology* 5 (01), 25–28

Westfall, P.H., Tobias, R.D. (Randall D., Wolfinger, R.D., and SAS Institute. (2011) *Multiple Comparisons and Multiple Tests Using SAS*. 2nd edn. SAS Pub

Wilcox, R.R. (2012) *Introduction to Robust Estimation and Hypothesis Testing*. 3rd edn. Academic Press

Winkel, R.R., von Euler-Chelpin, M., Nielsen, M., Petersen, K., Lillholm, M., Nielsen, M.B., Lynge, E., Uldall, W.Y., and Vejborg, I. (2016) 'Mammographic Density and

Structural Features Can Individually and Jointly Contribute to Breast Cancer Risk Assessment in Mammography Screening: A Case-Control Study.' *BMC Cancer* 16, 414

Wisdom, J.P., Cavaleri, M.A., Onwuegbuzie, A.J., and Green, C.A. (2012) 'Methodological Reporting in Qualitative, Quantitative, and Mixed Methods Health Services Research Articles'. *Health Services Research* 47 (2), 721–745

Wiswede, D., Koranyi, N., Müller, F., Langner, O., and Rothermund, K. (2013) 'Validating the Truth of Propositions: Behavioral and ERP Indicators of Truth Evaluation Processes'. *Social Cognitive and Affective Neuroscience* 8 (6), 647–653

Wolf, M., Krause, J., Carney, P.A., Bogart, A., and Kurvers, R.H.J.M. (2015) 'Collective Intelligence Meets Medical Decision-Making: The Collective Outperforms the Best Radiologist.' *PloS One* 10 (8), e0134269

Wolfe, J.N. (1976) 'Risk for Breast Cancer Development Determined by Mammographic Parenchymal Pattern.' *Cancer* 37 (5), 2486–92

Woolley, A.W., Aggarwal, I., and Malone, T.W. (2015) 'Collective Intelligence and Group Performance'. *Current Directions in Psychological Science* 24 (6), 420–424

Woolley, A.W., Chabris, C.F., Pentland, A., Hashmi, N., and Malone, T.W. (2010) 'Evidence for a Collective Intelligence Factor in the Performance of Human

Groups'. *Science* 330 (6004), 686–688

World Health Organisation (2018) *Cancer Tomorrow* [online] available from <www.https://gco.iarc.fr/tomorrow/graphic>

Worldwide Cancer Data | World Cancer Research Fund (n.d.) available from <https://www.wcrf.org/dietandcancer/cancer-trends/worldwide-cancer-data> [3 October 2018]

Yala, A., Lehman, C., Schuster, T., Portnoi, T., and Barzilay, R. (2019) 'A Deep Learning Mammography-Based Model for Improved Breast Cancer Risk Prediction'. *Radiology* 00 (0), 1–7

Yankaskas, B.C. (2004) *International Comparison of Performance Measures for Screening Mammography: Can It Be Done?* vol. 11

Yankaskas, B.C., Cleveland, R.J., Schell, M.J., and Kozar, R. (2001a) 'Association of Recall Rates with Sensitivity and Positive Predictive Values of Screening Mammography'. *American Journal of Roentgenology* 177 (3), 543–549

Yankaskas, B.C., Schell, M.J., Bird, R.E., and Desrochers, D.A. (2001b) 'Reassessment of Breast Cancers Missed During Routine Screening Mammography'. *American Journal of Roentgenology* 177 (3), 535–541

Yeung, N. and Summerfield, C. (2012) 'Metacognition in Human Decision-Making: Confidence and Error Monitoring'. *Philosophical Transactions of the Royal*

Society B: Biological Sciences 367 (1594), 1310–1321

Yoo, W., Mayberry, R., Bae, S., Singh, K., Peter He, Q., Lillard, J.W., and Jr. (2014) 'A Study of Effects of MultiCollinearity in the Multivariable Analysis.' *International Journal of Applied Science and Technology* 4 (5), 9–19

Zhang, X., Rice, M., Tworoger, S.S., Rosner, B.A., Eliassen, A.H., Tamimi, R.M., Joshi, A.D., Lindstrom, S., Qian, J., Colditz, G.A., Willett, W.C., Kraft, P., and Hankinson, S.E. (2018) 'Addition of a Polygenic Risk Score, Mammographic Density, and Endogenous Hormones to Existing Breast Cancer Risk Prediction Models: A Nested Case–Control Study'. *PLOS Medicine* 15 (9), e1002644

Ziegler, R.G., Hoover, R.N., Pike, M.C., Hildesheim, A., Nomura, A.M., West, D.W., Wu-Williams, A.H., Kolonel, L.N., Horn-Ross, P.L., Rosenthal, J.F., and Hyer, M.B. (1993) 'Migration Patterns and Breast Cancer Risk in Asian-American Women.' *Journal of the National Cancer Institute* 85 (22), 1819–1827

Zohar, D. and Hofmann, D.A. (2012) *Organizational Culture and Climate*. ed. by Kozlowski, J. Oxford University Press

Zwaan, L., Thijs, A., Wagner, C., van der Wal, G., and Timmermans, D.R.M. (2012) 'Relating Faults in Diagnostic Reasoning With Diagnostic Errors and Patient Harm'. *Academic Medicine* 87 (2), 149–156

Appendix 1. The NHS Breast Screening Programme Guidance on who can undertake arbitration

Public Health England

NHS Breast Screening Programme Guidance on who can undertake arbitration

1st edition August 2016

Public Health England leads the NHS Screening Programmes

About Public Health England

Public Health England exists to protect and improve the nation's health and wellbeing, and reduce health inequalities. It does this through world-class science, knowledge and intelligence, advocacy, partnerships and the delivery of specialist public health services. PHE is an operationally autonomous executive agency of the Department of Health.

Public Health England, Wellington House, 133-155 Waterloo Road, London SE1 8UG

Tel: 020 7654 8000 www.gov.uk/phe

Twitter: @PHE_uk Facebook: www.facebook.com/PublicHealthEngland

About PHE Screening

Screening identifies apparently healthy people who may be at increased risk of a disease or condition, enabling earlier treatment or better informed decisions. National population screening programmes are implemented in the NHS on the advice of the UK National Screening Committee (UK NSC), which makes independent, evidence-based recommendations to ministers in the four UK countries. The Screening Quality Assurance Service ensures programmes are safe and effective by checking that national standards are met. PHE leads the NHS Screening Programmes and hosts the UK NSC secretariat.

PHE Screening, Floor 2, Zone B, Skipton House, 80 London Road, London SE1 6LH www.gov.uk/topic/population-screening-programmes

Twitter: @PHE_Screening Blog: phescreening.blog.gov.uk

Prepared by: NHSBSP Radiology Co-ordinating GroupFor queries relating to this document, please contact: phe.screeninghelpdesk@nhs.net

Published August 2016

PHE publications gateway number: 2016

Contents

Executive summary

The process of arbitration occurs when 2 or more image readers do not reach consensus on the future management of the patient. This means that there are differing opinions as to whether a woman should be recalled to an assessment clinic or returned to routine screening in 3 years time, due to a perceived abnormality based on the interpretation of the images. The gold standard is for units to undertake arbitration of these cases.

Breast screening services undertake arbitration in different ways. Sometimes a third image reader will make a definitive decision to either recall a woman to assessment or return them to routine screening. Other services may convene a small group or panel of image readers to arbitrate on these cases. In situations where both readers have identified an abnormality, these cases may also be arbitrated/discussed according to the method of reading used by the service.

This document is based on expert opinion and gives guidance on who should undertake arbitration within a breast screening service

Introduction

This guidance is designed to assist the director of screening as to the suitability of a member of their team as a single arbitrator (or third reader) of screening mammograms for their programme. In those services where there is consensus, or team review of mammograms, this individual might also be the co-ordinator/lead of such a group, especially if new or inexperienced film reading staff are participating.

The arbitration process requires different competencies to those of film reading, especially decision making skills with good specificity. This skill set comes with experience, continuous feedback from clinical involvement and decision making in the assessment clinic, along with participation in audit, continuous professional development (CPD) and case review such as interval cancers. Clearly, the arbitrator cannot increase the sensitivity of the screen reading but can increase specificity and reduce the recall rate. These skills are not necessarily related to the profession of the arbitrator.

Recommended requirements for undertaking arbitration

Staff undertaking arbitration should:

be a fully qualified film reader meeting the appropriate standards including suitable training, reading $\geq$5000 films per year including 1500 first reads, 4000 screening mammograms

be an experienced film reader, >2 years in breast screening; if a new consultant radiologist, then full appropriate training must have been completed, with >5000 films read as a trainee, and ideally additional experience such as a breast fellowship post

participate fully in assessment clinics including decision making (working to consultant practitioner level)

regularly attend and participate at multi disciplinary team meetings (MDT). Minimum standard: “Colleagues involved in decision making and further diagnostic procedures (US and biopsy) should attend MDT meetings at which screening cases are discussed (twice per month on average) and/or should ensure that a formal process is in place for auditing their practice

and outcomes". NHSBSP Publication Number 49: Clinical Guidelines for Breast Cancer Screening Assessment 2010. Desirable standard: >20 per year

regularly audit and review personal and team results, with evidence of reflective learning, including: review of interval cancers, previously assessed intervals and screen detected cancers, and participation in Personal Performance in Mammographic Screening (PERFORMS)

participate in ongoing professional development and annual appraisal

The director of screening should agree that an individual is suitable for the role of arbitration and document this locally. The results of the individual and unit should be reviewed annually as part of local audit, clinical governance and the appraisal process. The Society and College of Radiographers (SCoR) can provide accreditation of advanced and consultant practitioners regarding the 4 pillars of practice which include: leadership, CPD and education, clinical practice and audit/research capabilities. Please see example link: https://www.sor.org/career-progression/consultants/consultant-practitioner-accreditation

Appendix 2. The TNM Classification for Breast Cancer

Primary tumour (T)	
TX	Primary tumour cannot be assessed
T0	No evidence of primary tumour
Tis	Carcinoma in situ
Tis (DCIS)	Ductal carcinoma in situ
Tis (LCIS)	Lobular carcinoma in situ
Tis (Paget)	Paget disease of the nipple NOT associated with invasive carcinoma and/or carcinoma in situ (DCIS and/or LCIS) in the underlying breast parenchyma. Carcinomas in the breast parenchyma associated with Paget disease are categorized based on the size and characteristics of the parenchymal disease, although the presence of Paget disease should still be noted
T1	Tumour ≤ 20 mm in greatest dimension
T1mi	Tumour ≤ 1 mm in greatest dimension
T1a	Tumour > 1 mm but ≤ 5 mm in greatest dimension
T1b	Tumour > 5 mm but ≤ 10 mm in greatest dimension
T1c	Tumour > 10 mm but ≤ 20 mm in greatest dimension
T2	Tumour > 20 mm but ≤ 50 mm in greatest dimension
T3	Tumour > 50 mm in greatest dimension
T4	Tumour of any size with direct extension to the chest wall and/or to the skin (ulceration or skin nodules)
T4a	Extension to chest wall, not including only pectoralis muscle adherence/invasion
T4b	Ulceration and/or ipsilateral satellite nodules and/or oedema (including peau d'orange) of the skin, which do not meet the criteria for inflammatory carcinoma

T4c	Both T4a and T4b
T4d	Inflammatory carcinoma
Regional lymph nodes (N)	
Clinical	
NX	Regional lymph nodes cannot be assessed (eg, previously removed)
N0	No regional lymph node metastasis
N1	Metastasis to movable ipsilateral level I, II axillary lymph node(s)
N2	Metastases in ipsilateral level I, II axillary lymph nodes that are clinically fixed or matted or in clinically detected* ipsilateral internal mammary nodes in the *absence* of clinically evident axillary lymph node metastasis
N2a	Metastases in ipsilateral level I, II axillary lymph nodes fixed to one another (matted) or to other structures
N2b	Metastases only in clinically detected* ipsilateral internal mammary nodes and in the *absence* of clinically evident level I, II axillary lymph node metastases
N3	Metastases in ipsilateral infraclavicular (level III axillary) lymph node(s), with or without level I, II axillary node involvement, or in clinically detected * ipsilateral internal mammary lymph node(s) and in the *presence* of clinically evident level I, II axillary lymph node metastasis; or metastasis in ipsilateral supraclavicular lymph node(s), with or without axillary or internal mammary lymph node involvement
N3a	Metastasis in ipsilateral infraclavicular lymph node(s)
N3b	Metastasis in ipsilateral internal mammary lymph node(s) and axillary lymph node(s)
N3c	Metastasis in ipsilateral supraclavicular lymph node(s)
*"Clinically detected" is defined as detected by imaging studies (excluding lymphoscintigraphy) or by clinical examination and having characteristics highly suspicious for malignancy or a presumed pathologic macrometastasis on the basis of fine-needle aspiration (FNA) biopsy with cytologic examination.	
Pathologic (pN)*	

pNX	Regional lymph nodes cannot be assessed (for example, previously removed, or not removed for pathologic study)
pN0	No regional lymph node metastasis identified histologically. *Note:* Isolated tumour cell clusters (ITCs) are defined as small clusters of cells ≤ 0.2 mm, or single tumour cells, or a cluster of < 200 cells in a single histologic cross-section; ITCs may be detected by routine histology or by immunohistochemical (IHC) methods; nodes containing only ITCs are excluded from the total positive node count for purposes of N classification but should be included in the total number of nodes evaluated
pN0(i-)	No regional lymph node metastases histologically, negative IHC
pN0(i+)	Malignant cells in regional lymph node(s) ≤ 0.2 mm (detected by hematoxylin-eosin [H&E] stain or IHC, including ITC)
pN0(mol-)	No regional lymph node metastases histologically, negative molecular findings (reverse transcriptase polymerase chain reaction [RT-PCR])
pN0(mol+)	Positive molecular findings (RT-PCR) but no regional lymph node metastases detected by histology or IHC
pN1	Micrometastases; or metastases in 1-3 axillary lymph nodes and/or in internal mammary nodes, with metastases detected by sentinel lymph node biopsy but not clinically detected†
pN1mi	Micrometastases (> 0.2 mm and/or > 200 cells, but none > 2.0 mm)
pN1a	Metastases in 1-3 axillary lymph nodes (at least 1 metastasis > 2.0 mm)
pN1b	Metastases in internal mammary nodes, with micrometastases or macrometastases detected by sentinel lymph node biopsy but not clinically detected†
pN1c	Metastases in 1-3 axillary lymph nodes and in internal mammary lymph nodes, with micrometastases or macrometastases detected by sentinel lymph node biopsy but not clinically detected†
pN2	Metastases in 4-9 axillary lymph nodes or in clinically detected‡ internal mammary lymph nodes in the absence of axillary lymph node metastases
pN2a	Metastases in 4-9 axillary lymph nodes (at least 1 tumour deposit > 2.0 mm)
pN2b	Metastases in clinically detected‡ internal mammary lymph nodes in the absence of axillary lymph node metastases

pN3	Metastases in ≥ 10 axillary lymph nodes; or in infraclavicular (level III axillary) lymph nodes; or in clinically detected‡ ipsilateral internal mammary lymph nodes in the presence of ≥ 1 positive level I, II axillary lymph nodes; or in > 3 axillary lymph nodes and in internal mammary lymph nodes, with micrometastases or macrometastases detected by sentinel lymph node biopsy but not clinically detected†; or in ipsilateral supraclavicular lymph nodes
pN3a	Metastases in ≥ 10 axillary lymph nodes (at least 1 tumour deposit > 2.0 mm); or metastases to the infraclavicular (level III axillary lymph) nodes
pN3b	Metastases in clinically detected‡ ipsilateral internal mammary lymph nodes in the presence of ≥ 1 positive axillary lymph nodes; or in > 3 axillary lymph nodes and in internal mammary lymph nodes, with micrometastases or macrometastases detected by sentinel lymph node biopsy but not clinically detected†
pN3c	Metastases in ipsilateral supraclavicular lymph nodes

*Classification is based on axillary lymph node dissection, with or without sentinel lymph node biopsy. Classification based solely on sentinel lymph node biopsy without subsequent axillary lymph node dissection is designated (sn) for "sentinel node"—for example, pN0(sn).

† "Not clinically detected" is defined as not detected by imaging studies (excluding lymphoscintigraphy) or not detected by clinical examination.

‡ "Clinically detected" is defined as detected by imaging studies (excluding lymphoscintigraphy) or by clinical examination and having characteristics highly suspicious for malignancy or a presumed pathologic macrometastasis on the basis of FNA biopsy with cytologic examination.

Distant metastasis (M)

M0	No clinical or radiographic evidence of distant metastasis
cM0(i+)	No clinical or radiographic evidence of distant metastases, but deposits of molecularly or microscopically detected tumour cells in circulating blood, bone marrow, or other non regional nodal tissue that are no larger than 0.2 mm in a patient without symptoms or signs of metastases
M1	Distant detectable metastases as determined by classic clinical and radiographic means and/or histologically proven > 0.2 mm

Appendix 3. Consolidated Standards for NHS Breast Screening Programme April 2017

BSP Standard 1	Inform: timely invitation letter sent to eligible women
Rationale	A key objective of the programme is to give women sufficient notice to be able to attend screening appointments allowing practical arrangements to be made to enable attendance and giving time for women to make an informed choice of whether to take up the offer of screening
Objective	To ensure that an appropriate timely and accessible screening invitation is sent to all eligible women
Criteria	The percentage of screening invitation letters giving at least two weeks' notice of the appointment date
Definitions	Numerator: Number of first offered invitations with ≥ 2 weeks before appointment date (50-70) Denominator: Total first offered invitations sent out to eligible screening population (50-70) (both within defined period expressed as a percentage) This excludes self and GP referrals
Performance thresholds	Acceptable ≥95% Achievable =100%
Mitigations/ qualifications	N/A
Reporting	*Reporting focus: screening service* *Data source: NBSS (to be developed)* *Responsible for submission: screening service* *Reporting period: Monthly (4 weeks in arrears)* *Quarterly (4 weeks in arrears*
BSP Standard 2	**Coverage: eligible population identified and invited**
Rationale	This standard is needed to ensure that the eligible population previously invited aged 53 to 70 has been adequately identified and invited by the screening programme
Objective	To maximise timely attendance within 36 months of screening in the eligible population
Criteria	The proportion of women eligible for screening who have had a test with a recorded result at least once in the previous 36 months
Definitions	Numerator: Number of eligible women aged 53-70 registered with a GP with a screening test result recorded in the past 36 months Denominator: Number of eligible women aged 53-70 registered with a GP (both within defined period expressed as a percentage) Women who are ineligible for screening due to having previously had a bi-lateral mastectomy, women who are ceased from the programme based on a "best interests" decision under the Mental Capacity Act 2005 or women who make an informed decision to remove themselves from the screening programme will be removed from the numerator and denominator There are a number of categories of women in the eligible age range who are not registered with a GP and subsequently not called for screening as they are not on the Breast Screening Select database. Screening units have a responsibility to maximise coverage of eligible women in their target population and should therefore be accessible to women in this category through self-referral and GP referral
Performance thresholds	Acceptable >70% Achievable >80%
Mitigations	All screening programmes should have the outcomes of women recorded and finalised within 6 months of their screening episode. If this is not done, it will adversely impact on rates of coverage. Screening services may have large numbers of women populating screening batches (for

	example with confederated GP groups) which may mean that closing screening episodes within the required 6-month interval is difficult. Some patient treatment regimes may expand beyond 6 months (egg, where neo-adjuvant therapies administered) which will mean some patient episodes will not be closed within 6 months. If screening programmes have any screening slippage (all women not invited within 36 months of their previous screen), it will adversely impact on rates of coverage. Further, it will invalidate many performance measures which are based on a 36-month screening interval.
Reporting	*Reporting focus: Local Authority* *Data source: Breast Screening Select* *Responsible for submission: Exeter, NHS Digital* *Monthly and annual reporting schedules (6 months in arrears)*
BSP Standard 3	**Maximising effectiveness of the screening programme: Uptake rates**
Rationale	The expected effectiveness of the breast screening programme in reducing breast cancer mortality requires uptake to be maximised.
Objective	To maximise uptake in the eligible population who are fully informed and wish to participate in the screening programme
Criteria	The percentage of eligible women invited who attend for screening
Definitions	Numerator: Total eligible women attending screening (within 6 months of data of first offered appointment Denominator; Total eligible women with date of first offered appointment within the period (both within defined period expressed as a percentage) The uptake standard counts appointments not women. If a woman is invited more than once during a year, she will have more than one screening episode counted during the period. Second timed appointments are not counted as a second screening episode
Performance thresholds	Acceptable >70% Achievable >80%
Mitigations	N/A
Reporting	*Reporting focus: screening service* *Data source: NBSS (KC62 report: Tables A-C2 aged 50-70)* *Responsible for submission: screening service* Data on this indicator will only be accurate 6 months after the end of the reporting period. Care should be taken when reviewing provisional quarterly data due to the proportion of open episodes where women have yet to attend an appointment. Quarterly (provisional data produced 4 weeks in arrears) Annual (definitive data produced 6 months in arrears)
Equity impact	Hard to reach and vulnerable groups may be the least likely to attend. Programmes should work to ensure that their local population demographics are known and that all women have equal opportunity to make an informed choice and have access to the service via local health promotion initiatives. Analysis of uptake rates by GP screening practice are recommended.
BSP Standard 4	**Uptake: Maintaining screening round length**
Rationale	Delivering and maintaining round length is important to help achieve the desired mortality reduction. This is achieved by detecting incident screen cancers as early as possible and minimising interval cancers (cancers presenting in between screening episodes) and reducing the negative consequences of inviting women too frequently
Objective	To ensure that women are recalled for screening at 36 month intervals
Criteria	The percentage of eligible women whose date of first offered appointment is within 36 months of their previous screen. Women being screening for the first time will not be included in screening round length statistics
Definitions	Numerator: Number of eligible women aged 50-70yrs with date of first offered

	appointment within 36 months of their previous screen within the report period Denominator: Total number of eligible women (50-70 yrs.) screened (both within defined period expressed as a percentage) This excludes self and GP referrals
Performance thresholds	Acceptable ≥90% Achievable 100%
Mitigations	Breast Screening select was introduced in July 2016. This has replaced NHAIS to facilitate call and recall. The transition away from NHAIS has resulted in the removal of area code as a method to select screening batches and GP out code has taken its place (this is available on the spine). This could cause screening slippage at some services as the cohort definition has now been changed. This effect could be felt for the 36 months following implementation.
Reporting	*Reporting focus: screening service* *Data source: NBSS* *Responsible for submission: screening service* Monthly and quarterly (produced 4 weeks in arrears)
BSP Standard 5	**Test and minimising harm: Repeat examination rate**
Rationale	There is a balance between radiation dose and image quality. Services should aim to deliver the optimum image quality with as low a radiation dose as possible. To ensure good quality practice the number of repeat examinations is monitored.
Objective	To minimise the number of women undergoing repeat examinations to minimise anxiety and exposure to radiation
Criteria	The proportion of repeat examinations (due to technical recalls or technical repeats) by service (also recommended by practitioner)
Definitions	Numerator: Total number of women requiring repeat examinations Denominator: Total number of women attending screening (both within defined period expressed as a percentage) The measure is calculated with the trainee film readers Repeat mammography rates may be higher for trainee mammographers or assistant practitioners than trained staff. It is advisable to calculate the rates both including and excluding trainees.
Performance thresholds	Acceptable <3% Achievable <2%
Mitigations	N/A
Reporting	*Reporting focus: screening service* *Data source: NBSS* *Responsible for submission: screening service* Monthly and quarterly (produced 4 weeks in arrears)
BSP Standard 6	**Minimising harm: recording appropriate radiation dose**
Rationale	To ensure that the radiation dose from the mammograms used for screening and assessment is as low as possible and to ensure the minimum harm to women from the radiation used, whilst providing sufficient image quality for cancer detection.
Objective	To limit the amount of radiation dose to the glandular tissues of the breast from mammograms
Criteria	Mean glandular dose (MGD) per view for a standard breast in clinical settings
Definitions	The method of estimating the mean glandular dose to a standard breast using a 45mm thick Perspex (PMMA) phantom is described in "Commissioning and routine testing of full field digital mammography systems" (NHSBSP Equipment Report 1303) https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/442720/nhsbsp-equipment-report-1303.pdf
Performance	Acceptable ≤2.5mGy

<table>
<tr><td>thresholds</td><td></td></tr>
<tr><td>Mitigations</td><td>N/a</td></tr>
<tr><td>Reporting</td><td>Reporting focus: screening service digital mammography (2-D) equipment
Data source: screening service physics survey report
Responsible for submission: screening unit physics service
The MGD to the standard breast for each mammography system used in the NHSBSP is measured by a medical physics service routinely every 6 months and after major changes to the equipment and reported through the Quality Control system.</td></tr>
<tr><td colspan="2"></td></tr>
<tr><td>BSP Standard 7</td><td>Minimising harm and diagnosis: image quality</td></tr>
<tr><td>Rationale</td><td>This standard is to ensure the technical image quality of mammograms used for screening and assessment is sufficient to achieve the objectives of cancer detection</td></tr>
<tr><td>Objective</td><td>To maximise the numbers of cancers detected</td></tr>
<tr><td>Criteria</td><td>Threshold gold thickness measured using the CDMAM test object</td></tr>
<tr><td>Definitions</td><td>The method of measuring threshold gold thickness is described in “Commissioning and routine testing of full field digital mammography systems” (NHSBSP Equipment Report 1303).
https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/442720/nhsbsp-equipment-report-1303.pdf
Software is provided by the NHSBSP to automate the analysis of CDMAM images for 0.1 to 1.0 mm detail sizes.</td></tr>
<tr><td>Performance thresholds</td><td>
<table>
<tr><td></td><td colspan="2">Threshold gold thickness (μm) *</td></tr>
<tr><td>Diameter of detail (mm)</td><td>Minimum acceptable value</td><td>Achievable Value</td></tr>
<tr><td>1</td><td>≤0.091</td><td>≤0.056</td></tr>
<tr><td>0.5</td><td>≤0.150</td><td>≤0.103</td></tr>
<tr><td>0.25</td><td>≤0.352</td><td>≤0.244</td></tr>
<tr><td>0.1</td><td>≤1.68</td><td>≤1.10</td></tr>
</table>
* Lower values of threshold gold thickness indicate better image quality
</td></tr>
<tr><td>Mitigations</td><td>If a measurement appears to be above the standard, the CDMAM test object should be considered as there is some variability in measurement between test objects</td></tr>
<tr><td>Reporting</td><td>Reporting focus: screening service digital mammography (2-D) equipment
Data source: NBSS
Responsible for submission: screening service
The image quality for each mammography system used in the NHSBSP is measured by a medical physics service every 6 months and reported through the Quality Control system</td></tr>
<tr><td colspan="2"></td></tr>
<tr><td>BSP Standard 8</td><td>Minimising harm: receipt of screening results</td></tr>
<tr><td>Rationale</td><td>It is essential that women receive the results of screening in a timely manner to ensure those who require further tests and those who do not are informed at the earliest opportunity</td></tr>
<tr><td>Objective</td><td>To minimise anxiety for women who are awaiting the results of screening</td></tr>
<tr><td>Criteria</td><td>The proportion of women who are sent their result within two weeks of an adequate screen</td></tr>
<tr><td>Definitions</td><td>Numerator: Total adequately screened women sent results within 2 weeks
Denominator: Total adequately screened women sent results
(both within defined period expressed as a percentage)</td></tr>
</table>

Performance Thresholds	Acceptable ≥95% Achievable 100%
Mitigations	N/a
Reporting	*Reporting focus: screening service* *Data source: NBSS* *Responsible for submission: screening service* *Monthly and quarterly (produced 4 weeks in arrears)*
BSP Standard 9	**Minimising harm: referral to assessment rates**
Rationale	To encourage high specificity and should be examined together with cancer detection rates to ensure that both screening specificity and sensitivity are maximised. Those responsible for interpreting the images from breast screening need to ensure that they are recalling the right women with abnormalities which require further investigation whilst not recalling too many women where no abnormalities are subsequently found.
Objective	To minimise the number of women screened who are referred for further tests whilst trying to minimise false negative rates
Criteria	The proportion of eligible women with a technically adequate screen who are referred for assessment
Definitions	Numerator: Number of adequately screened women referred for assessment Denominator: Total number of eligible women with a technically adequate screen (both within defined period expressed as a percentage)
Performance Thresholds	Acceptable < 10% (prevalent screen) < 7% (incident screen) Achievable <7% (prevalent screen), <5% (incident screen)
Mitigations	Screening services may not always seek to reduce recall rates depending on levels of cancer detection. Where particularly high cancer detection rates are found it may not always be feasible to reduce referral for assessment rates. New image readers are expected to have higher rates of referral on average than experienced readers.
Reporting	*Reporting focus: screening service* *Data source: NBSS (KC62 report)* *Responsible for submission: screening service* *Quarterly (6 weeks in arrears), and annually (definitive data 6 months in arrears)* Prevalent screen includes women aged 45-52 (from KC62 Table A) Incident screen includes women aged 50-70 (from KC62 Table C1)
BSP Standard 10	**Minimising harm: short-term recall rates**
Rationale	Every effort should be made to obtain a definitive diagnosis at initial assessment and short-term recall should be used only in exceptional circumstances and with informed consent, as it is associated with significant anxiety
Objective	To minimise the number of women who are recalled for further tests one year after previous assessment
Criteria	The percentage of women screened who are placed on short term recall
Definitions	Numerator: Number of eligible women screened given short-term recall appointment Denominator: Number of eligible women adequately screened (both within defined period expressed as a percentage)
Performance Thresholds	Acceptable <0.25% Achievable <0.12% There are rare occurrences when a short-term recall may be justified but women should not receive more than one short-term recall outcome within a normal three yearly screening episode
Mitigations	N/a
Reporting	*Reporting focus: screening service* *Data source: NBSS (KC62, table T, aged 50-70)*

	Responsible for submission: screening service *Quarterly (6 weeks in arrears), and annually (definitive data 6 months in arrears)*
BSP Standard 11	**Minimising harm: time to first offered appointment for assessment**
Rationale	It is important to minimise anxiety in women who need to attend for further screening tests to obtain a definitive malignant, benign or normal diagnosis
Objective	To minimise the interval from the screening mammogram to assessment
Criteria	The percentage of women who are offered an appointment at an assessment centre within three weeks of attendance for the screening mammogram
Definitions	Numerator: Number of eligible women whose first offered appointment for assessment is within 3 weeks of an initial adequate screen Denominator: Number of eligible women referred for assessment (both within defined period expressed as a percentage)
Performance thresholds	Acceptable >98% Achievable 100%
Mitigations	N/a
Reporting	*Reporting focus: screening service* *Data source: NBSS* *Responsible for submission: screening service* *Monthly and quarterly (6 weeks in areas)*
BSP standard 12	**Minimising harm: number of assessment visits to obtain a definitive diagnosis**
Rationale	It is important to reduce anxiety in women by aiming to minimise the number of assessment visits required in order to obtain a definitive diagnosis. An early non-operative diagnosis of malignancy is highly desirable as it allows informed pre-treatment counselling of the patient and facilitates one-stage treatment thus ensuring that anxiety is minimised.
Objective	The number of diagnostic assessment visits needed to achieve a definitive outcome should be as low as possible.
Criteria	The minimum standard is that 95% of women should require no more than 3 separate visits for diagnostic assessment (including visits to receive results). The number of visits will depend on the structure of the assessment process; however, no more than 2 needle biopsy procedures carried out on separate occasions should normally be needed to achieve a non-operative diagnosis.
Definitions	Numerator: Number of women with ≤3 visits for diagnostic assessment and results appointments Denominator: Number of eligible women attending assessment (both within defined period expressed as a percentage)
Performance thresholds	Acceptable ≥95%
Mitigations	In some circumstances, repeated visits may be necessary where difficult to diagnose lesions are found to be multi-focal or the MDT requires further investigations to be undertaken. Some services may not have the resources to allow all investigations to be undertaken in one visit. This may lead to more than two visits for further diagnostic tests on occasion.
Reporting	*Reporting focus: screening service* *Data source: NBSS* *Responsible for submission: screening service* Annually as part of the Association of Breast Surgeons audit
BSP Standard 13	**Minimising harm: benign biopsies rates**
Rationale	To minimise unnecessary surgery as the number of open surgical biopsies performed as a result of screening that prove to be benign should be as low as possible given high rates of non-operative diagnosis in the Programme

Objective	To minimise the number of unnecessary operative procedures
Criteria	To minimise the rate of surgical benign biopsies
Definitions	Numerator: Number of surgical biopsies with a benign or normal histological outcome Denominator: Number of eligible women with a technically adequate screen (both within defined period expressed as a rate per 1000 screened)
Performance thresholds	Acceptable < 1.5/1000 (prevalent screen) < 1.0/1000 (incident screen) Achievable <1/1000 (prevalent screen), <0.75/1000 (incident screen)
Mitigations	Lack of availability or access to vacuum assisted biopsy could impact on the number of women referred onwards to open surgical biopsy.
Reporting	*Reporting focus: screening service* *Data source: NBSS (KC62)* *Responsible for submission: screening service* 6 monthly (provisional data), annually (definitive data) 6 months in arrears Prevalent screen includes women aged 45-52 (from KC62 Table A) Incident screen includes women aged 50-70 (from KC62 Table C1)
BSP Standard 14	**Diagnose: rates of non-operative diagnosis**
Rationale	It is important to minimise the number of operative procedures necessary and to enable treatment planning in advance of surgery
Objective	To ensure that the majority of cancers, both palpable and impalpable receive a non-operative tissue diagnosis of cancer
Criteria	The number of women who have a non-operative diagnosis of cancer by needle histology or cytology after a maximum of two visits expressed as a proportion of all women screened diagnosed with breast cancer
Definitions	Numerator: Number of women with non-operative diagnosis (within 2 visits to assessment) Denominator: Number of women diagnosed with breast cancer (both within defined period expressed as a percentage)
Performance thresholds	Acceptable ≥90% (invasive disease), >=85% (non-invasive disease) Achievable ≥ 95% (invasive disease), >= 90% (non-invasive disease)
Mitigations	Services should report non-invasive diagnosis rates both with and without lobular carcinoma in situ (LCIS) as this will impact on non-operative diagnosis rates achievable.
Reporting	*Reporting focus: screening service* *Data source: NBSS (KC62, table T, 50-70) and ABS audit for information on with/without LCIS* *Responsible for submission: screening service* *Bi-annually (provisional data), annually (6 months in arrears-definitive data)*
BSP Standard 15	**Diagnose: age standardised detection ratios (SDRs for invasive cancers)**
Rationale	It is important to compare cancer detection between screening services with differing mean ages of screening populations, as the age of women screened is a major determinant of cancer detection rates. This is corrected for by using a standardised detection rate which allows the observed invasive cancers to be compared to the expected number of invasive cancers, given the age distribution of the population screened
Objective	To maximise the numbers of invasive cancers detected
Criteria	The SDR is the ratio of the observed number of invasive cancers to the expected number in the eligible population invited and screened
Definitions	Numerator: Number of women with invasive cancer in eligible women screened Denominator: The expected number of invasive cancers in eligible women screened (both within defined period) The expected number of cancers is based on applying criteria from the Swedish Two Counties randomised control trial which is used as a comparator of performance
Performance thresholds	Acceptable 1.00 Achievable 1.40
Mitigations	The reporting breast screening service may refer women for treatment to alternative

	providers. Sometimes it can be difficult to obtain the pathology and treatment details accurately for entry onto NBSS which may mean that cancers may be under-reported by the host service where the woman was initially screened.
Reporting	*Reporting focus: screening service* *Data source: NBSS* *Responsible for submission: screening service* Bi-annually (provisional data), annually (6 months in arrears-definitive data) Prevalent screen includes women aged 45-70 (from KC62 Table A + B) Incident screen includes women aged 50-70 (from KC62 Table C1)
BSP Standard 16	**Diagnose: small cancer age standardised detection ratios (invasive cancers)**
Rationale	To achieve a significant reduction in breast cancer mortality it is of significant importance that small invasive breast cancers (< 15 mm diameter) are detected.
Objective	To maximise the numbers of small cancers detected
Criteria	The standardised detection ration (SDR) is the ratio of the observed number of invasive cancers to the expected number in the eligible population invited and screened. Small cancers (<15mm in diameter) should be 55% of the expected overall number of invasive cancers.
Definitions	Numerator: Number of women with invasive cancer diagnosed <15mm in diameter Denominator: The expected number of invasive cancers diagnosed <15mm in diameter (both within defined period)
Performance thresholds	Acceptable 1.00 Achievable 1.40
Mitigations	The size distribution of all invasive cancers should be examined to establish whether there is any "rounding up" of cancers measuring between 14mm and 15mm by pathologists. If this is shown, it may reduce the numbers of small cancers detected Host screening services may refer women for treatment to alternative providers. Sometimes it can be difficult to obtain the pathology and treatment details accurately for entry onto NBSS which may mean that cancers may be under-reported by the host service where the woman was initially screened
Reporting	*Reporting focus: screening service* *Data source: NBSS (KC62,)* *Responsible for submission: screening service* Bi-annually (provisional data), annually (6 months in arrears-definitive data) All screens aged 45-70 (from KC62 Tables A+B+C1)
BSP Standard 17	**Diagnose: non-invasive cancer detection rates**
Rationale	Detection of non-invasive cancer at screening (predominantly ductal carcinoma in situ (DCIS), particularly high-grade types, is assumed to be a factor contributing to long-term reduction in mortality although no firm scientific evidence currently exists to confirm this. The majority of DCIS detected at screening is of the high-risk type. It is believed to be good practice to detect and treat DCIS
Objective	To ensure that the rate of non-invasive cancer is maximised (particularly high grade disease)
Criteria	The rate of cancers detected that are non-invasive (*in situ)* carcinoma
Definitions	Numerator: Number of women with non and micro-invasive cancers Denominator: Number of eligible women with a technically adequate screen (both within defined period expressed as a rate per 1000 screened)
Performance thresholds	Acceptable ≥0.5/1000 (prevalent screen), >=0.6/1000 (incident screen) Achievable n/a Some experts have argued that detection of this stage of breast carcinoma may represent over diagnosis (detecting disease which would never become clinically apparent or threaten life) and causes anxiety and physical harm (unnecessary surgery). Others suggest that detection of DCIS is important because they believe that it is a precursor of invasive carcinoma. Until the Sloane Study can give definitive evidence, Programme advice is to maximise detection of non-invasive cancer (particularly high

	grade disease).
Mitigations	N/a
Reporting	*Reporting focus: screening service* *Data source: NBSS (KC62)* *Responsible for submission: screening service* Bi-annually (provisional data), annually (6 months in arrears-definitive data) Prevalent screen includes women aged 45-70 (from KC62 Table A) Incident screen includes women aged 50-70 (from KC62 Table C1)
BSP Standard 18	**Diagnose: staging of the axilla**
Rationale	It is important to allow planning for appropriate patient management at the earliest opportunity if suspected or diagnosed cancer has spread to the axilla.
Objective	To ensure adequate staging of the axilla in patients with invasive breast cancer.
Criteria	Patients treated surgically for early invasive breast cancer should have an axillary staging procedure carried out if metastatic nodal metastasis is not confirmed non-operatively
Definitions	Numerator: Number of women with invasive breast cancer with an axillary staging procedure Denominator: Number of women with invasive breast cancer (both within defined period expressed as a percentage)
Performance thresholds	Acceptable: >90% Achievable 100%
Mitigations	N/a
Reporting	*Reporting focus: screening service* *Data source: NBSS* *Responsible for submission: screening service* Annually all ages as part of the Association of Breast Surgeons audit
BSP Standard 19	**Outcomes: rates of interval cancers**
Rationale	Cancers that are detected between screens (Interval Cancers) decrease the likelihood of reducing mortality in the eligible screening population.
Objective	To minimise the number of interval cancers presenting between screening episodes
Criteria	The number of interval cancers per 1000 women screened
Definitions	Numerator: Number of women eligible for screening presenting with interval cancers within 36 months of a previous screen Denominator: Total number of eligible women screened (Number of women screened within a screening year and interval cancers arising within 36 months of the specified period expressed as a rate per 1000 screened)
Performance thresholds	Acceptable: <0.65/1000 diagnosed <12 months of the previous screen <1.40/1000 diagnosed between 12 and <24 months of the previous screen <1.65/1000 diagnosed between 24 and <36 months of the previous screen Achievable: n/a Analysis of interval cancer data should take place at screening service level aggregating several years' performance, as the number of interval cancers occurring in individual screening units each year is relatively small and analysis of them is likely to be meaningful only when several years' data are available. Interval cancers should be examined alongside other screening data (such as SDRs) when considering the performance of a breast screening programme as failure to achieve interval cancer targets may coincide with high rates of cancer detection and may reflect higher than expected rates of cancer prevalence in the underlying population or failure to meet screening round length target
Mitigations	N/a

Reporting	*Reporting focus: screening service* *Data source: NBSS & Screening Histories Information Management system (SHIM)* *Responsible for submission: screening service* Annual audit for women aged 47-73 at screening

Appendix 4. Table 10. Studies exploring the association between recall rates and performance measures

1*. Yankaskas et al. *(2001) USA				
Research question/aim	**Study design**	**Method** **Data Collection** **Sample size**	**Data Analysis/Metrics**	**Main findings/results**
To measure the effect on sensitivity and positive predictive value as recall rates increase.	Prospective design	31 Community-based mammography facilities January 1994 - June 1998	Reduced monotonic regression analysis to model PPV and sensitivity rates as functions of the recall rate Linear regression analysis to examine the association of recall rates with sensitivity and PPV (adjustments for relevant Covariates). overall recall rate = 6.3%. Highest recall= 13.6% and the lowest 2.4% Sensitivity was inversely related to recall rates for age. The same inverse relationship (decrease in recall rate and increase in sensitivity) was seen for a decrease in breast	Practices with recall rates between 4.9% and 5.5% achieved the best trade-off of sensitivity and PPV. 1-year cancer detection rate was 3.5 per 1000 mammograms Recall rates decreased with increasing practice volume

			density, for a personal history of breast cancer, for a history of breast surgery, and for the presence of breast symptoms. PPV was inversely related to recall rate for age, an increase in breast density and an increase in the time elapsed since previous mammography.	
2*. Gur et al. *(2004) USA				
Research question/aim	**Study design**	**Method** **Data Collection** **Sample size**	**Data Analysis/Metrics**	**Main findings/results**
Investigate the correlation between recall rates and CDR	Retrospective Design	10 Radiologists 2000-2002	Parametric Pearson (r) and nonparametric Spearman (rho) correlation coefficients.	A wide range of recall rates (range, 7.7–17.2%) and detection rates (range,2.6 –5.4 per 1000 mammograms) Linear fit between recall rates and CDR ($p < 0.05$) Higher recall rates =higher detection rates. Increase in detection rate extended beyond the recommended practice guideline of 10%.

3*. Yankaskas et al. *(2004) USA				
Research question/aim	**Study design**	**Method** **Data Collection** **Sample size**	**Data Analysis/Metrics**	**Main findings/results**
To assess the comparability of recall rates and PPV internationally.	Prospective design 3 phase project Phase 1-an assessment form to obtain descriptive data on how screening was performed and what specific data related to recall was collected. 2 - data collected for calculation of recall rates and PPV 3 - more targeted and strata-specific data for calculation of recall rates, PPV and cancer detection.	22 countries 1997 - 2002.	Multivariate analysis Wide variation in recall rates – 1.4-15.1% USA= highest recall rates (15%) and lowest in the Netherlands (1.4%). PPV rates range 5–37.5%	Increasing recall rate= decreasing PPV CDR showed less variation than recall and PPV rates, CDR per 1000 ranged from 3.9–10.6. Netherlands = low recall rate + high PPV= CDR of 5.3/1000. No direct relationship of recall to cancer incidence

4*. Otten et al. *(2005) Netherlands				
Research question/aim	**Study design**	**Method** **Data Collection** **Sample size**	**Data Analysis/Metrics**	**Main findings/results**
To estimate the effect of changes in recall rate on earlier detection of cancer	Retrospective	495 sets of screen-negative mammograms. 250 controls 245 subsequently diagnosed with breast cancer 1997 – 1999 15 Radiologists – various countries 5 regional screening organisations	Mean detection sensitivities for different false-positive rates calculated with a linear mixed model. Localization-response receiver operating characteristic (LROC)	Breast cancer can be detected earlier by especially for recall rates of 1% – 4%. Above 4% CDR levels off with a disproportionate increase of false-positive rates.
5*. Schell et al. *(2007) USA				
Research question/aim	**Study design**	**Method** **Data Collection** **Sample size**	**Data Analysis/Metrics**	**Main findings/results**
Identify target recall rates for screening. Investigate how sensitivity shifts with recall rate.	Retrospective design	1996-2001 6 sites	1. Isotonic regression analysis 2. Reduced monotonic regression 3. Reduced monotonic regression model 4. Concave fit	Recommend recall rates of 10.0% for prevalent and 6.7% for incident screens based on additional workups per additional cancers detected.

6*. Grabler et al. *(2017) USA				
Research question/aim	**Study design**	**Method Data Collection Sample size**	**Data Analysis/Metrics**	**Main findings/results**
Identify the optimal screening recall rate range based on CDR's	Retrospective design	2007-2012	Linear regression with bootstrap bias correction to assess changes in CDR with increases in the recall rate.	"sweet spot" for optimal cancer detection is in the recall range 12-14%. Incremental benefit above this is relatively small. A recall rate less than 10% may be too low.
7*. Mullen et al. *(2017) USA				
Research question/aim	**Study design**	**Method Data Collection Sample size**	**Data Analysis/Metrics**	**Main findings/results**
Determine the impact of interventions designed to reduce screening mammography recall rates on screening performance metric	Retrospective and prospective – 2 stage design	2012 - 2016 Assessed baseline performance for FFDM and DBT 1st intervention - readers reviewed their own recalls, outcomes of diagnostic evaluation and biopsy. 2nd -intervention was consensus double reading of all recalls, 3rd reader arbitrator	Pearson two-tailed chi squared tests to determine the effect of each intervention on recall rates and performance metrics compared to the baseline. Data stratified by age groups to assess whether trends were uniformly seen across the study population	The baseline recall rate, cancer detection rate, and PPV1 were 11.1%, 3.8/1000, and 3.4%, respectively, for FFDM, and 7.6%,4.8/1000, and 6.0%, respectively, for DBT. Recall rates decreased significantly to 9.2% for FFDM and to 6.6% for DBT after the 1st intervention as well as to 9.9% for FFDM after the second intervention. PPV1 increased significantly to 5.7% for FFDM and to 9.0% for DBT after the second intervention. Cancer detection rate did not significantly change with either intervention. An average of 2.3 minutes was spent consulting for each recall.

		10 radiologists, all breast imaging specialists		
8. **Taylor-Phillips et al. (2017) UK**				
Research question/aim	**Study design**	**Method Data Collection Sample size**	**Data Analysis/Metrics**	**Main findings/results**
Investigate the effect of double readings by a second radiologist on recall rates, cancer detection, and characteristics of cancers	Retrospective analysis	805 206 women evaluated through screening and diagnostic test results - 1 year of routine data from 33 English breast screening centres	the test for equality of proportions, the x2 test for independence, and the t test.	Double reading with arbitration reduces recall and increases cancer detection compared with single reading. Cancers detected only by the second reader were smaller, of lower grade, and had less nodal involvement.

Appendix 5. Health Research Authority (HRA) Approval

Content removed on data protection grounds

Content removed on data protection grounds

Content removed on data protection grounds

Content removed on data protection grounds

Appendix A - List of Documents

The final document set assessed and approved by HRA Approval is listed below.

Document	*Version*	*Date*
Confirmation of any other Regulatory Approvals (e.g. NIGB) and all correspondence [Coventry University Phase 1 Ethics Certificate]	V1	
Confirmation of any other Regulatory Approvals (e.g. NIGB) and all correspondence [Coventry University Phase 2 ethics certificate]	V1	
Confirmation of any other Regulatory Approvals (e.g. NIGB) and all correspondence [CU review document]	V1	26 May 2017
Contract/Study Agreement [Indemnity]	V1	26 May 2017
Copies of advertisement materials for research participants [BSBR newsletter advertisement]	V2	21 March 2017
Copies of advertisement materials for research participants [Journal Advert]	V3	20 April 2017
Copies of advertisement materials for research participants [Twitter Advert]	V1	21 March 2017
Evidence of Sponsor insurance or indemnity (non NHS Sponsors only) [Liability]	V1	26 May 2017
HRA Schedule of Events [hra-schedule-events-exceltemplate-3 (8) Interviews (Dated)]	1.0	19 July 2017
HRA Schedule of Events [hra-schedule-events-exceltemplate-3 (Survey) Dated]	1.0	19 July 2017
HRA Statement of Activities [statement-activities-telephone interview Dated]	1.0	19 July 2017
HRA Statement of Activities [statement-activities-survey Dated]	1.0	19 July 2017
Interview schedules or topic guides for participants [Telephone interview schedule]	V2	02 May 2017
IRAS Application Form [IRAS_Form_07062017]		07 June 2017
Letter from funder [CU letter]	V1	26 May 2017
Letter from sponsor [Sponsor letter]	V1	26 May 2017
Letters of invitation to participant [Email invitation for telephone interviews]	V1	28 April 2017

Letters of invitation to participant [Cover e-mail]	V1	21 April 2017
Letters of invitation to participant [Survey Introduction letter]	V1	21 April 2017
Non-validated questionnaire [Film reader survey]	V10	26 May 2017
Non-validated questionnaire [Director Survey]	V10	26 May 2017
Participant information sheet (PIS) [PIS for telephone interviews]	V2	04 November 2016
Participant information sheet (PIS) [PIS Film reader survey]	V5	04 November 2016
Participant information sheet (PIS) [PIS Director survey]	V5	04 November 2016
Participant information sheet (PIS) [Director PIS IRAS]	5	04 November 2016
Participant information sheet (PIS) [Film Reader PIS IRAS]	5	04 November 2016
Participant information sheet (PIS) [PIS_for_semistructured_interviews_IRAS]	2	04 November 2016
Research protocol or project proposal [Study protocol]	V1	26 April 2017
Summary CV for Chief Investigator (CI) [CV]	V1	26 May 2017
Summary CV for student [Summary CV]	V1	26 May 2017
Summary CV for supervisor (student research) [Supervisor CV]	V1	01 February 2017
Summary, synopsis or diagram (flowchart) of protocol in non technical language [Flow chart of study]	V1	19 April 2017

Appendix B - Summary of HRA Assessment

Content removed on data protection grounds

Content removed on data protection grounds

Content removed on data protection grounds

Content removed on data protection grounds

Content removed on data protection grounds

Content removed on data protection grounds

Director of Breast Screening Survey

Study Information

IRAS Project ID: **228030**

"A survey to explore reporting and arbitration/consensus processes within Breast Screening Units in England."

You are invited to take part in the above-named research study. Please read the following information sheet that explains about the study before you decide if you wish to participate.
No individual or unit will be identified in the report.

My professional background is a Consultant radiographer in breast imaging. This study is part of a PhD supported by a Coventry University studentship.

The study involves an **online survey** containing questions relating to reporting and arbitration/consensus review meetings in breast screening. There are six sections relating to the workforce, reporting practices, arbitration and consensus practice, and Public Health England arbitration guidance. **It is anticipated that the survey should take less than 20 minutes to complete.**

Purpose of the Study

The purpose of this study is to map current reporting and arbitration practices within breast screening units in England. Questionnaire information will complement published (KC62) data to establish if certain strategies function better in differing units. The survey aims to investigate the use of radiographers as 3rd reader arbitrators (or lead of consensus meetings) and to identify any associated barriers.

The breast screening reporter survey also aims to determine if radiographers participating in the survey meet the recommended requirements for undertaking arbitration as specified within the recent Public Health England guidance

(www.gov.uk/government/publications/breast-screening-arbitration-guidance). **The survey**
results will be kept confidential, PHE and trusts will not have access to any individual data.

Who is doing the study?

This study is being conducted by Lisa Hackney, Consultant Radiographer and PhD student at Coventry University. Lisa Hackney is being supervised by Professor Derek Renshaw, Professor Ala Szczepura, Dr Louise Moody, and Becky Whiteman and has received a Coventry University Studentship to support her PhD study and this research project.

Who is being asked to participate?

Directors of breast screening units within England are being invited to participate in the study.

A supplementary survey is available for Radiologists, Radiographers, Breast Clinicians and other staff who are currently reporting in breast screening services.

Do I have to take part?

Your participation in this study is entirely voluntary, and you do not have to give reasons for not participating. However, your support in providing the information will be greatly appreciated as this will portray a comprehensive representation of current practice in England.

What will be involved if I take part in this study?

You will be asked to complete an online questionnaire. We anticipate that the survey should take less than 20 minutes to complete. The survey includes factors relating to the workforce, reporting practices, arbitration and consensus practice, and Public Health England arbitration guidance.

How will consent be obtained?

After reading the online information page and before commencing the questionnaire you will be required to answer a mandatory question **"Do you agree to take part in this study", Yes or No**.

If you agree to take part, you will automatically be directed to the questionnaire.

If you decline to take part you will automatically be directed to the end, therefore being unable to view or complete the questionnaire.

What are the risks associated with this project?

We believe that there are no risks associated with completing or not completing the questionnaire.

What are the benefits of taking part?

There are no immediate benefits to you as an individual if you take part in the study. However, there is a lack of published literature relating to arbitration and consensus processes within breast screening. Within England, there is a wide variation in recall rates (particularly for prevalent screens) with a number of units not achieving the NHSBSP standards for assessment recall (minimum of <10% with a target of <7%) and arbitration/consensus can be integral in achieving this. Therefore, you will be contributing to data that will help us to evaluate and analyse variations in practice. It is essential to gather information from as many units as possible to portray a representative and comprehensive 'snap shot' of existing and planned practice.

Can I withdraw from the study at any time?

If you decide not to complete the survey the partial information you have entered will not be used. You do not have to give reasons for non-completion. If you fully complete and submit the survey, it will not be possible to extract your data. Therefore, withdrawal after submission is not possible.

Will the information I give be kept confidential?

Study data will be handled in accordance with the Data Protection Act 1998 and Coventry University safeguarding data policy. Findings will **not** be identifiable **by you or your unit** but may be grouped via QA regions. All data will be retained on password protected computers and encrypted data sticks. All study data will be destroyed three years after the PhD has been completed.

What will happen to the results of the study?

Please note that no breast unit response will be identifiable in any final report. Once the survey is complete, we will send you a summary of responses to the questions posed. Your responses go directly to Lisa Hackney at Coventry

University, and no individual or unit data will be shared with anyone (including PHE). The results of the study will be submitted to Coventry University as part of Lisa Hackney's PhD award, and will also be utilised to form the basis of papers and posters submitted to national and international conferences and peer-reviewed journals. A summary report of overall results will be produced for PHE.

Who has reviewed this study?

Ethical approval for this study has been granted by Coventry University Research Ethics and Governance Committee *(reference: P45921)* on *6th February 2017,* and the Health Research Authority (IRAS: 228030) on 31st July 2017. The study questionnaire has been reviewed and piloted by experienced academics and clinical professionals.

What if I want to complain?

Content removed on data protection grounds

Thank you for taking the time to consider this request.

Consent

Do you agree to take part in this study? *Required*

- Yes
- No

This questionnaire has been **divided into six sections** relating to reporting and arbitration/ consensus processes within breast screening. Dependent upon responses, some sections/questions will not require completion. If you are **unable** to complete elements of the questionnaire, **please forward to a member of your breast radiology team** who will be able to provide the relevant information. It is anticipated that the survey should take less than 20 minutes to complete.

For this survey, arbitration is classified as either by a solitary 3rd reader who makes the final decision on their own

Consensus is defined as a pair or group (2 or more individuals) decision-making process. Group members discuss and agree to support a decision even if not the "preferent" of each individual.

Section 1- Workforce

Section 2 – Reporting Practice

Section 3 - Current Arbitration/Consensus practice

Section 4 – Scheduling

Section 5 - Consensus Practice

Section 6 – Guidance/Implementation

This online document can be accessed at various time intervals prior to submission; previously entered data will be automatically saved between pages.

Please complete this questionnaire by **13th September 2017**

Unit

Please provide the full name of your breast screening unit. **We will not** publish this information. The information will be used to group unit responses into geographical regions and to complement published (KC62) data. Findings will **not** be identifiable **by you or your unit** *Required*

Section 1: Workforce

How many years of experience do you have in breast screen reading?
Required

Please enter a whole number (integer).

Please state the number (Headcount) of professionals currently undertaking breast **screen reading** in your unit. (*Please enter the number within each category)*

	Number
Radiologist	
Advanced Practitioner Radiographer	
Consultant Radiographer	
Breast Clinician	
Locum Radiologist	
Locum Advanced Practitioner Radiographer	
Locum Consultant Radiographer	
Locum Breast Clinician	
Other	

If you selected other, please specify the role

In your unit, please specify which professionals **currently** make the final decision on their own (solitary 3rd reader) on arbitration cases? (*Please tick for all that apply) Required*

- ☐ Not applicable - 3rd reader arbitration not used
- ☐ Radiologist
- ☐ Advanced Practitioner Radiographer
- ☐ Consultant Radiographer
- ☐ Breast Clinician
- ☐ Locum Radiologist
- ☐ Locum Advanced Practitioner Radiographer
- ☐ Locum Consultant Radiographer
- ☐ Locum Breast Clinician
- ☐ Other

If you selected Other, please specify:

In your unit, please specify which professionals **currently** take the lead in consensus meetings (responsible for the final report on NBSS)? *Please tick for all that apply Required*

- ☐ Not applicable - consensus not used
- ☐ Not applicable - no lead
- ☐ Radiologist
- ☐ Advanced Practitioner Radiographer

- [ ] Consultant Radiographer
- [ ] Breast Clinician
- [ ] Locum Radiologist
- [ ] Locum Advanced Practitioner
- [ ] Radiographer
- [ ] Locum Consultant Radiographer
- [ ] Locum Breast Clinician
- [ ] Other

If you selected Other, please specify:

Please state who takes responsibility for the final report on NBSS

Section 2: Reporting Practice

Does your unit restrict reporters reading together in any way? (e.g. based on experience, profession, recall rates, cancer detection rates)? *Required*

- Yes
- No

If yes, please explain

Does your unit use double radiographer reporting (both first and second readers are radiographers)? *Required*

- Yes, routine practice
- Yes, but not routinely
- No

Please indicate which of the following represents reporting practice at your unit. (*Please select only one option) Required*

- Blinded double reading (the second reader cannot see the first reader's decision on the computer software or assessment paperwork)

- Blinded double reading (the second reader cannot see the first reader’s decision on the computer software but can by looking at the assessment paperwork)
- Non-blinded double reading (first reader’s decision is visible on screen)
- Other

If you selected Other, please specify:

Please include any additional comments on reading practice's (optional)

Section 3: Current Arbitration/Consensus practice

Prevalent screening - which cases does your unit arbitrate or review at consensus meetings? (*Please select only one option) Required*

- All recalled cases (concordant and discordant recalls)
- Disagreement only cases (discordant recalls)
- Other

If you selected Other, please specify:

Incident screening - which cases does your unit arbitrate or review at consensus meetings? *(Please select only one option) Required*

- All recalled cases (concordant and discordant recalls)
- Disagreement only cases (discordant recalls)
- Other

If you selected Other, please specify:

Is there any difference in the cases you arbitrate or send for consensus review if the reporting was undertaken by **two radiographers**, as opposed to one (or both) reporter/s being a **Radiologist**? *Required*

- ○ Yes
- ○ No
- ○ Not applicable (unit does not use double radiographer reporting)

Please specify

Please select which strategy your unit uses to resolve **discordant prevalent** screening cases. (*Please select all that apply) Required*

- ☐ Automatically recall if 1 reader specifies
- ☐ 3rd Reader Arbitrator has the final decision
- ☐ Consensus pair (2 readers different from the original reporters)
- ☐ Consensus pair (2 readers - 1 of which may be an original reporter)
- ☐ Consensus group (3 or more readers different from the original reporters)
- ☐ Consensus group (3 or more readers including one or both of the original reporters)
- ☐ Other

If you selected Other, please specify:

If you selected multiple options - please explain what determines which approach is used

Please select which strategy your unit uses to resolve **discordant incident** screening cases? (Please select all that apply) *Required*

- ☐ Automatically recall if 1 reader specifies
- ☐ 3rd Reader Arbitrator has the final decision
- ☐ Consensus pair (2 readers different from the original reporters)
- ☐ Consensus pair (2 readers - 1 of which may be an original reporter)
- ☐ Consensus group (3 or more readers different from the original reporters)
- ☐ Consensus group (3 or more readers including one or both of the original reporters)
- ☐ Other

If you selected Other, please specify:

If you selected multiple options - please explain what determines which approach is used

In your opinion, what were the main reasons your current strategies (to resolve discordant prevalent and incident cases) were implemented? *Required*

Please describe any data that was used to support the system your unit implemented to resolve arbitration/consensus cases (if none, please state none)?

Does your unit adhere to written protocols (SOPS) for resolving arbitration/consensus cases? *Required*

- Yes
- No
- I don't know

Has your unit sent cases externally (to another breast screening service) to be arbitrated/consensus reviewed? *Required*

- Yes
- No

Please explain why

Has your unit failed screen to routine recall and/or screen to assessment targets as a result of cases awaiting arbitration/consensus review? *Required*

- Yes -only in 2017
- Yes – on occasion within the last 5 years
- No – never
- Other

If you selected

Section 4: Scheduling

Does your unit undertake single 3rd reader arbitration or consensus review meetings on scheduled day/s of the week? *Required*

- Yes
- No

Please explain what determines when (day(s)/time of day) single 3rd reader arbitration or consensus review is undertaken

In an average working week, how many arbitration sessions and/or consensus meetings does **your unit** have? *Required*

In an average consensus meeting, how many staff participate? If single 3rd reader arbitration **only** performed - please enter 1 *Required*

In an **average working week**, please estimate the amount of time (minutes) your unit dedicates to arbitration and/or consensus meetings *Required*

In an average working week, please estimate the number of cases arbitrated and/or reviewed at consensus meetings *Required*

Please include any additional comments regarding the scheduling of arbitration or consensus meetings (optional)

Section 5: Consensus

This section is only relevant if your unit performs some form of consensus review (for concordant and/or discordant cases). Therefore, if you automatically recall if 1 reader specifies or undertake solitary 3rd reader arbitration you will be directed to the final section

Does your unit undertake any form of consensus (pair or group) review for recalled cases (concordant and/or discordant)? *Required*

- Yes
- No

Does your unit specify a minimum (Quorum) membership (grade and/or a number of staff) required for the consensus review meeting to go ahead? *Required*

- Yes
- No

Please state what the quorum membership requirements are

At consensus meetings which strategy does your unit use to make the final decision on a case? (*Please select only 1 option) Required*

- Recall if any individual specifies
- Majority decision (equal skills assumed)
- Decision weighted by experience
- Decision weighted by profession
- Other

If you selected Other, please specify:

Within your unit please rate the following statements regarding the consensus group. (*Please choose only one option per statement) Required*

Please select at least 15 answer(s).

	Strongly Agree	Agree	Neither agree nor disagree	Disagree	Strongly Disagree
Membership of the consensus group changes frequently so there isn't a set team					
The consensus group has the right "mix" of staff—a group of people who bring different clinical perspectives and experiences to the discussion					
There is a real desire among team members in the consensus group to work collaboratively					
Each group member shares accountability for consensus group decisions and outcomes					

Consensus meetings provide an open, comfortable, safe place to discuss cases					

Please don't select more than 1 answer(s) per row.

When team members disagree, all points of view are considered before deciding on the final outcome	☐	☐	☐	☐	☐
During the meeting, team members ask for and give each other constructive feedback.	☐	☐	☐	☐	☐
Our team has mechanisms in place to monitor consensus outcomes	☐	☐	☐	☐	☐
Within the consensus group, we are able to work through differences of opinion without damaging relationships.	☐	☐	☐	☐	☐
Members of the consensus team depend on each other for their special knowledge and expertise	☐	☐	☐	☐	☐
Members of the consensus group show respect for each other's roles and expertise	☐	☐	☐	☐	☐

The way the consensus group members interact improves the quality of patient care	☐	☐	☐	☐	☐
I feel integral to the consensus group	☐	☐	☐	☐	☐
Experience excellent teamwork with the members of the consensus group	☐	☐	☐	☐	☐
Consensus meetings provide an opportunity for educational learning from cases	☐	☐	☐	☐	☐

Please include any additional comments on consensus practice (optional)

Section 6: Guidance and Implementation

Has your unit ever utilised radiographers as single 3rd reader arbitrators making the final decision (responsible for the final report on NBSS) on cases? *Required*

- ○ Yes
- ○ No
- ○ I don't know

What grades were the radiographers?

- ☐ Advanced Practitioner Radiographer (qualified film reader)
- ☐ Consultant Radiographer
- ☐ Other

If you selected Other, please specify:

If within the last 5 years please state the time period/s (month/year) when this was used (otherwise, state more than 5 years ago).

Please state what criteria your unit used to determine a radiographer within your team was suitable to perform single 3rd reader arbitration, making the final decision on arbitration cases?

Has your unit ever utilised radiographers as the lead for consensus meetings (responsible for the final report on NBSS)? *Required*

- ○ Yes
- ○ No
- ○ I don't know

What grades were the radiographers?

- ☐ Advanced Practitioner Radiographer (qualified film reader)
- ☐ Consultant Radiographer
- ☐ Other

If you selected Other, please specify:

If within the last 5 years, please state the time period/s (month/year) when this was used (otherwise state more than 5 years ago)

Please state what criteria your unit used to determine a radiographer within your team was suitable to take the lead of consensus review meetings (responsible for the final report on NBSS)?

Will (or has) the Public Health England 2016 screening guidance on arbitration change(d) practice in your unit? *Required*

- Yes
- No
- I don't know
- I don't know about the guidance (https://www.gov.uk/government/publications/breast-screening-arbitration-guidance)

Please specify - and what grades of staff have you (will you be) delegated (delegating) to?

In your opinion, please rate the following statements as to why solitary 3rd reader radiographer arbitration (radiographer lead of consensus) may not be implemented in your unit. *Required*

Please don't select more than 1 answer(s) per row.

Please select at least 11 answer(s).

	Strongly Agree	Agree	Neither agree nor disagree	Disagree	Strongly Disagree
The unit has a sufficient number of Radiologists to undertake solitary 3rd reader arbitration (or lead consensus meetings)					
No Radiographers in the unit meet the recommended requirements within the PHE guidance					
There is no organisational support to delegate arbitration (lead consensus) to radiographers					
Concern that recall rates may increase with radiographer arbitration					

Concern that cancer detection rates may decrease with radiographer arbitration					

Radiographers in the unit do not have the leadership skills to co-ordinate a consensus meeting	☐	☐	☐	☐	☐
Radiographers in the unit do not want to undertake this role	☐	☐	☐	☐	☐
The organisational culture means it takes time to embed change	☐	☐	☐	☐	☐
Individual Radiologists are resistant to change	☐	☐	☐	☐	☐
Radiologists feel threatened by task shifting (radiographer arbitration/lead of consensus)	☐	☐	☐	☐	☐
There is no good reason not to implement radiographer arbitration/lead of consensus	☐	☐	☐	☐	☐

Other constraints/challenges to implementing radiographer arbitration/lead of consensus-please specify (optional)

Final Survey Comments

Please provide any additional feedback on your views/experiences of arbitration or consensus that has not already been included in this questionnaire. (Optional)

Thank you very much for participating

Thank you very much for your valuable contribution to this study. We may wish to contact you for further details following this questionnaire. Please state whether you are willing to participate in a telephone interview (all data would be **confidential and anonymous**)? *Required*

- Yes
- No

Please provide your contact details - e-mail address (this will not be retained after the study is complete)

Please enter a valid email address.

For further information, please contact Lisa Hackney

Content removed on data protection grounds

(https://coventry.onlinesurveys.ac.uk/breast-screening-reporter-survey**) to all professional groups that report breast screening within your unit (Radiologists, Radiographers, Breast Clinicians etc.)**

Breast Screening Reporter Survey

Study Information

IRAS Project ID: **228030**

"A survey to explore reporting and arbitration/consensus processes within Breast Screening Units in England."

You are invited to take part in the above-named research study. Please read the following information sheet that explains about the study before you decide if you wish to participate.
No individual or unit will be identified in the report.

My professional background is a Consultant radiographer in breast imaging. This study is part of a PhD supported by a Coventry University studentship.

The study involves an **online survey** containing questions relating to reporting and arbitration/consensus review meetings in breast screening. There are six sections relating to the workforce, reporting practices, arbitration and consensus practice, and Public Health England arbitration guidance. **It is anticipated that the survey should take less than 20 minutes to complete.**

Purpose of the Study

The purpose of this study is to map current reporting and arbitration practices within breast screening units in England. Questionnaire information will complement published (KC62) data to establish if certain strategies function better in differing units. The survey aims to investigate the use of radiographers as 3rd reader arbitrators (or lead of consensus meetings) and to identify any associated barriers.

The breast screening reporter survey also aims to determine if radiographers participating in the survey meet the recommended requirements for undertaking arbitration as specified within the recent Public Health England guidance

(www.gov.uk/government/publications/breast-screening-arbitration-guidance). **The survey**
results will be kept confidential, PHE and trusts will not have access to any individual data.

Who is doing the study?

This study is being conducted by Lisa Hackney, Consultant Radiographer and PhD student at Coventry University. Lisa Hackney is being supervised by Professor Derek Renshaw, Professor Ala Szczepura, Dr Louise Moody, and Becky Whiteman and has received a Coventry University Studentship to support her PhD study and this research project.

Who is being asked to participate?

Directors of breast screening units within England are being invited to participate in the study.

A supplementary survey is available for Radiologists, Radiographers, Breast Clinicians and other staff who are currently reporting in breast screening services.

Do I have to take part?

Your participation in this study is entirely voluntary, and you do not have to give reasons for not participating. However, your support in providing the information will be greatly appreciated as this will portray a comprehensive representation of current practice in England.

What will be involved if I take part in this study?

You will be asked to complete an online questionnaire. We anticipate that the survey should take less than 20 minutes to complete. The survey includes factors relating to the workforce, reporting practices, arbitration and consensus practice, and Public Health England arbitration guidance.

How will consent be obtained?

After reading the online information page and before commencing the questionnaire you will be required to answer a mandatory question **"Do you agree to take part in this study", Yes or No**.

If you agree to take part, you will automatically be directed to the questionnaire.

What are the risks associated with this project?

We believe that there are no risks associated with completing or not completing the questionnaire.

What are the benefits of taking part?

There are no immediate benefits to you as an individual if you take part in the study. However, there is a lack of published literature relating to arbitration and consensus processes within breast screening. Within England, there is a wide variation in recall rates (particularly for prevalent screens) with a number of units not achieving the NHSBSP standards for assessment recall (minimum of <10% with a target of <7%) and arbitration/consensus can be integral in achieving this. Therefore, you will be contributing to data that will help us to evaluate and analyse variations in practice. It is essential to gather information from as many units as possible to portray a representative and comprehensive 'snap shot' of existing and planned practice.

Can I withdraw from the study at any time?

If you decide not to complete the survey the partial information you have entered will not be used. You do not have to give reasons for non-completion. If you fully complete and submit the survey, it will not be possible to extract your data. Therefore, withdrawal after submission is not possible.

Will the information I give be kept confidential?

Study data will be handled in accordance with the Data Protection Act 1998 and Coventry University safeguarding data policy. Findings will **not** be identifiable **by you or your unit** but may be grouped via QA regions. All data will be retained on password protected computers and encrypted data sticks. All study data will be destroyed three years after the PhD has been completed

What will happen to the results of the study?

Please note that no breast unit response will be identifiable in any final report. Your responses go directly to Lisa Hackney at Coventry University, and no individual or unit data will be shared with anyone (including PHE). The results of the study will be submitted to Coventry University as part of Lisa Hackney's PhD award, and will also be utilised to form the basis of papers and posters submitted to national and international

conferences and peer-reviewed journals. A summary report of overall results will be produced for PHE.

Who has reviewed this study?

Ethical approval for this study has been granted by Coventry University Research Ethics and Governance Committee *(reference: P45921)* on *6th February 2017*, and the Health Research Authority (IRAS: 228030) on 31st July 2017. The study questionnaire has been reviewed and piloted by experienced academics and clinical professionals.

What if I want to complain?

Content removed on data protection grounds

Thank you for taking the time to consider this request

Consent

Do you agree to take part in this study? *Required*

- Yes
- No

This questionnaire has been **divided into six sections** relating to reporting and arbitration/ consensus processes within breast screening. Dependent upon responses, some sections/questions will not require completion. It is anticipated that the survey should take less than 20 minutes to complete.

For this survey, arbitration is classified as either a solitary 3rd reader who makes the final decision on their own.

Consensus is defined as a pair or group (2 or more individuals) decision-making process. Group members discuss and agree to support a decision even if not the "preferent" of each individual.

Section 1- Workforce

Section 2 – Reporting Practice

Section 3 - Current Arbitration/Consensus practice

Section 4 – Scheduling

Section 5 - Consensus Practice

Section 6 – Guidance/Implementation

This online document can be accessed at various time intervals prior to submission; previously entered data will be automatically saved between pages.

Please complete this questionnaire by **13th September 2017**

Unit

Please provide the full name of your breast screening unit. **We will not** publish this information. The information will be used to group unit responses into geographical regions and to complement published (KC62) data. Findings will **not** be identifiable **by you or your unit** *Required*

Section 1: Workforce

How many years of experience do **you** have in breast **screen reading**?

Please enter a whole number (integer).

In your unit, please specify which professionals **currently** make the final decision on their own (solitary 3rd reader) on arbitration cases? *Please tick for all that apply*

Required

- ☐ Not applicable - 3rd reader arbitration not used
- ☐ Radiologist
- ☐ Advanced Practitioner Radiographer
- ☐ Consultant Radiographer
- ☐ Breast Clinician
- ☐ Locum Radiologist
- ☐ Locum Advanced Practitioner Radiographer
- ☐ Locum Consultant Radiographer
- ☐ Locum Breast Clinician
- ☐ Other

If you selected Other, please specify:

In your unit, please specify which professionals **currently** take the lead in consensus meetings (responsible for the final report on NBSS)? *Please tick for all that apply* *Required*

- [ ] Not applicable - consensus not used
- [ ] Not applicable - no lead
- [ ] Radiologist
- [ ] Advanced Practitioner Radiographer
- [ ] Consultant Radiographer
- [ ] Breast Clinician
- [ ] Locum Radiologist
- [ ] Locum Advanced Practitioner Radiographer
- [ ] Locum Consultant Radiographer
- [ ] Locum Breast Clinician
- [ ] Other

If you selected Other, please specify:

Please state who takes responsibility for the final report on NBSS

Section 2: Reporting Practice

Does your unit restrict reporters reading together in any way? (e.g. based on experience, profession, recall rates, cancer detection rates)? *Required*

- Yes
- No

If yes, please explain

Does your unit use double radiographer reporting (both first and second readers are radiographers)? *Required*

- Yes, routine practice
- Yes, but not routinely
- No

Please indicate which of the following represents reporting practice at your unit. (*Please select only one option)* *Required*

- Blinded double reading (the second reader cannot see the first reader's decision on the computer software or assessment paperwork)

- Blinded double reading (the second reader cannot see the first reader's decision on the computer software but can by looking at the assessment paperwork)
- Non-blinded double reading (first reader's decision is visible on screen)
- Other

If you selected Other, please specify:

Please include any additional comments on reading practice's (optional)

Section 3: Current Arbitration/Consensus practice

Prevalent screening - which cases does your unit arbitrate or review at consensus meetings? (*Please select only one option)* *Required*

- ○ All recalled cases (concordant and discordant recalls)
- ○ Disagreement only cases (discordant recalls)
- ○ Other

If you selected Other, please specify:

Incident screening - which cases does your unit arbitrate or review at consensus meetings? *(Please select only one option)* *Required*

- ○ All recalled cases (concordant and discordant recalls)
- ○ Disagreement only cases (discordant recalls)
- ○ Other

If you selected Other, please specify:

Is there any difference in the cases you arbitrate or send for consensus review if the reporting was undertaken by **two radiographers,** as opposed to one (or both) reporter/s being a **Radiologist**? *Required*

- ○ Yes
- ○ No
- ○ Not applicable (unit does not use double radiographer reporting)

Please specify

Please select which strategy your unit uses to resolve **discordant prevalent** screening cases. (*Please tick all that apply)* *Required*

- ☐ Automatically recall if 1 reader specifies
- ☐ 3rd Reader Arbitrator has the final decision
- ☐ Consensus pair (2 readers different from the original reporters)
- ☐ Consensus pair (2 readers - 1 of which may be an original reporter)
- ☐ Consensus group (3 or more readers different from the original reporters)
- ☐ Consensus group (3 or more readers including one or both of the original reporters)
- ☐ Other

If you selected Other, please specify:

If you selected multiple options - please explain what determines which approach is used

Please select which strategy your unit uses to resolve **discordant incident** screening cases? (Please tick for all that apply) *Required*

- ☐ Automatically recall if 1 reader specifies
- ☐ 3rd Reader Arbitrator has the final decision
- ☐ Consensus pair (2 readers different from the original reporters)
- ☐ Consensus pair (2 readers - 1 of which may be an original reporter)
- ☐ Consensus group (3 or more readers different from the original reporters)
- ☐ Consensus group (3 or more readers including one or both of the original reporters)
- ☐ Other

If you selected Other, please specify:

If you selected multiple options - please explain what determines which approach is used

In your opinion, what were the main reasons your current strategies (to resolve discordant prevalent and incident cases) were implemented? *Required*

Does your unit adhere to written protocols (SOPS) for resolving arbitration/consensus cases? *Required*

- Yes
- No
- Don't know

Section 4: Scheduling

Does your unit undertake single 3rd reader arbitration **or** consensus review meetings on scheduled day/s of the week? *Required*

- Yes
- No

Please explain what determines when (day(s)/time of day) single 3rd reader arbitration or consensus review is undertaken

In an average working week, how many arbitration sessions and/or consensus meetings does your unit have? *Required*

In an average consensus meeting, how many staff participate? If single 3rd reader arbitration only performed - please enter 1 *Required*

In an average working week, please estimate the amount of time (**minutes)** **your unit** dedicates to arbitration **and/or** consensus meetings *Required*

In an **average working week**, please estimate the number of cases arbitrated **and/or** reviewed at consensus meetings *Required*

Please include any additional comments regarding the scheduling of arbitration or consensus meetings (optional)

Section 5: Consensus

This section is only relevant if your unit performs some form of consensus review (for concordant and/or discordant cases). Therefore, if you automatically recall if 1 reader specifies or undertake solitary 3rd reader arbitration you will be directed to the final section

Does your unit undertake any form of consensus (pair or group) review for recalled cases (concordant and/or discordant)? *Required*

- Yes
- No

Does your unit specify a minimum (Quorum) membership (grade and/or a number of staff) required for the consensus review meeting to go ahead? *Required*

- Yes
- No

Please state what the quorum membership requirements are

At consensus meetings which strategy does your unit use to make the final decision on a case? (*Please select only 1 option)* *Required*

- Recall if any individual specifies
- Majority decision (equal skills assumed)
- Decision weighted by experience
- Decision weighted by profession
- Other

If you selected Other, please specify:

Within your unit please rate the following statements regarding the consensus group. (*Please choose only one option per statement)* *Required*

Please don't select more than 1 answer(s) per row.

Please select at least 15 answer(s).

	Strongly Agree	Agree	Neither agree nor disagree	Disagree	Stron Disag
Membership of the consensus group changes frequently so there isn't a set team					
The consensus group has the right "mix" of staff—a group of people who bring different clinical perspectives and experiences to the discussion					
There is a real desire among team members in the consensus group to work collaboratively					
Each group member shares accountability for consensus group decisions and outcomes					

Consensus meetings provide an open, comfortable, safe place to discuss cases					

When team members disagree, all points of view are considered before deciding on the final outcome	☐	☐	☐	☐	☐
During the meeting, team members ask for and give each other constructive feedback.	☐	☐	☐	☐	☐
Our team has mechanisms in place to monitor consensus outcomes	☐	☐	☐	☐	☐
Within the consensus group, we are able to work through differences of opinion without damaging relationships.	☐	☐	☐	☐	☐
Members of the consensus team depend on each other for their special knowledge and expertise	☐	☐	☐	☐	☐
Members of the consensus group show respect for each other's roles and expertise	☐	☐	☐	☐	☐

The way the consensus group members interact improves the quality of patient care	☐	☐	☐	☐	☐
I feel integral to the consensus group	☐	☐	☐	☐	☐
Experience excellent teamwork with the members of the consensus group	☐	☐	☐	☐	☐
Consensus meetings provide an opportunity for educational learning from cases	☐	☐	☐	☐	☐

Please include any additional comments on consensus practice (optional)

Professional role

What is your current professional role? (*Please tick only one option*) *Required*

- ○ Radiologist
- ○ Advanced Practitioner Radiographer
- ○ Consultant Radiographer
- ○ Breast Clinician
- ○ Locum Radiologist
- ○ Locum Advanced Practitioner Radiographer
- ○ Locum Consultant Radiographer
- ○ Locum Breast Clinician
- ○ Other

If you selected Other, please specify:

Section 6: Guidance and Implementation

Has your unit ever utilised radiographers as single 3rd reader arbitrators making the final decision (responsible for the final report on NBSS) on cases? *Required*

- ○ Yes
- ○ No
- ○ I don't know

What grades were the radiographers?

- ☐ Advanced practitioners (qualified film reader)
- ☐ Consultant Radiographer
- ☐ Other

If you selected Other, please specify:

If within the last 5 years, please state the time period/s (month/year) when this was used (otherwise state more than 5 years ago)

Has your unit ever utilised radiographers as the lead for consensus meetings (responsible for the final report on NBSS)? *Required*

- ○ Yes
- ○ No
- ○ I don't know

What grades were the radiographers?

- ☐ Advanced Practitioner Radiographer (qualified film reader)
- ☐ Consultant Radiographer
- ☐ Other

If you selected Other, please specify:

If within the last 5 years, please state the time period/s (month/year) when this was used (otherwise state more than 5 years ago)

Do you (personally) **currently** undertake single 3rd reader of arbitration cases, making the final decision on your own? *Required*

- Yes
- No

Please state what criteria was used to determine your suitability to perform single 3rd reader arbitration

Do you (personally) **currently** take the lead of consensus review meetings, (responsible for the final report on NBSS)? *Required*

- Yes
- No

Please state what criteria was used to determine your suitability to lead consensus meetings

Will (or has) the Public Health England 2016 screening guidance on arbitration change(d) practice in your unit? *Required*

- Yes
- No
- I don't know
- I don't know about the guidance (https://www.gov.uk/government/publications/breast-screening-arbitration-guidance)

Please specify - what grades of staff has (is) the director delegated (planning on delegating) to?

In your opinion, please rate the following statements as to why solitary 3rd reader radiographer arbitration (radiographer lead of consensus) may not be implemented in your unit. *Required*

Please don't select more than 1 answer(s) per row.

Please select at least 11 answer(s).

	Strongly Agree	Agree	Neither agree nor disagree	Disagree	Strongly Disagree
The unit has a sufficient number of Radiologists to undertake solitary 3rd reader arbitration (or lead consensus meetings)					
No Radiographers in the unit meet the recommended requirements within the PHE guidance					
There is no organisational support to delegate arbitration (lead consensus) to radiographers					
Concern that recall rates may increase with radiographer arbitration					

Concern that cancer detection rates may decrease with radiographer arbitration					

Radiographers in the unit do not have the leadership skills to co-ordinate a consensus meeting	☐	☐	☐	☐	☐
Radiographers in the unit do not want to undertake this role	☐	☐	☐	☐	☐
The organisational culture means it takes time to embed change	☐	☐	☐	☐	☐
Individual Radiologists are resistant to change	☐	☐	☐	☐	☐
Radiologists feel threatened by task shifting (radiographer arbitration/lead of consensus)	☐	☐	☐	☐	☐
There is no good reason not to implement radiographer arbitration/lead of consensus	☐	☐	☐	☐	☐

Other constraints/challenges to implementing radiographer arbitration/lead of consensus-please specify (optional)

Please rate how often you have…… *Required*

Please don't select more than 1 answer(s) per row.

Please select at least 4 answer(s).

	Never	Occasionally	(intermittent years)	Annually
Read > 5000 films per annum	☐	☐	☐	☐
First read 1500 screening films per annum	☐	☐	☐	☐
Participated in PERFORMS	☐	☐	☐	☐
An annual appraisal /personal development review	☐	☐	☐	☐

Please state if you meet the following criteria…… *Required*

Please don't select more than 1 answer(s) per row.

Please select at least 5 answer(s).

	Yes	No	working towards
Undertake autonomous decision making in assessment clinics	☐	☐	☐
Contribute to decision-making (not just attend) at MDT meetings	☐	☐	☐
Undertake regular audit and review of personal reading results	☐	☐	☐
Undertake regular audit and review of team results	☐	☐	☐
Evidence reflective learning from review of interval cancers, previously assessed intervals and screen-detected cancers	☐	☐	☐

Are you accredited with The Society and College of Radiographers (SCoR) as an advanced or consultant practitioner? (*Please select only one option)*
Required

- Yes, currently
- No, but previously have been accredited
- Never been accredited

Final Survey Comments

Please provide any additional feedback on your views/experiences of arbitration or consensus that has not already been included in this questionnaire. (Optional)

Thank you very much for participating

Thank you very much for your valuable contribution to this study. We may wish to contact you for further details following this questionnaire. Please state whether you are willing to participate in a telephone interview (all data would be confidential and anonymous)? *Required*

- Yes
- No

Please provide your contact details - e-mail address (this will not be retained after the study is complete)

Please enter a valid email address.

For further information, please contact Lisa Hackney

Content removed on data protection grounds

Appendix 8. Table 28-1A Free Text Survey Comments

6.9.4.1: Blinded vs non-blinded reading

"The second read is supposed to be blind, but the paperwork can easily be seen – it is stamped 'RECALL' by the first and this can be seen"	*Film reader 10 – Consultant radiographer*
"Blinding on software optional but not enforced – variable practice"	*Film reader 15 – Consultant radiographer*
"Personally, I don't look at the decision of another reader if available on screen until I have made my own regardless of reading first or second"	*Director 11*
"I personally prefer to use blind data entry on NBSS but do not think my colleagues do"	*Director 3*
"We use in-house screening forms that are designed to convey information clearly and tend to note our thoughts when we have spent some time looking at an area, whether we recall it or note. I believe this helps to share experience between readers"	*Director 2*

Table 28-1B Free Text Survey Comments 6.9.4.2 Reporting restrictions

"Radiographers have to read with a radiologist and cannot read together"	*Director 10*
"At least one of the two readers is always a radiologist"	*Director 17*
"Radiographers – advanced and consultant are not allowed to read together"	*Director 27*
"Most advanced practitioners only first read"	*Film reader 2 –*

	Advanced Practitioner
"Radiographer film readers third read as a trainee until they have completed 5000 reads. After that, they first read only and not with another Radiographer film reader"	*Film reader 49 – Advanced Practitioner*
"Newly qualified readers only read first, not second for their first 5,000 sets of mammograms"	*Director 9*
"Newly qualified reader (< 2 years' experience) can only read with a more experienced (> 2 years' experience) reader"	*Director 24*
"We do try and pair certain readers, but it is not always possible to adhere to this based on film reader data sent to clinical director"	*Film reader 55 – Consultant Radiographer*
"Advised certain pairs not to read, e.g. if both have a high recall rate"	*Film reader 41 - Radiologist*

Table 28-1C Free Text Survey Comments 6.9.5.1 Strategies to resolve discordant cases

"Availability of staff. Consensus meeting performed with whichever film readers available"	*Director 23*
"Third reader arbitration used only rarely when there is not enough staff"	*Director 12*
"Depends on staffing levels leave etc. as to which is possible without delaying woman's result/recall"	*Film reader 8 – Consultant Radiographer*
"*Depends who/how many people are available. Try not to use original reporters, but sometimes they're all that are*	*Film reader 2 –*

available"	*Advanced Practitioner*
"Our routine practice is to arbitrate, but if staffing allows, we take the opportunity to consensus with whoever is available"	*Film reader 35 – Consultant Radiographer*
"If a lot to be reviewed we will illuminate concordant results only if one reader was a Radiologist"	*Film reader 14 – Advanced Practitioner*

Table 28-1D Free Text Survey Comments 6.9.5.2 Rationale/data to support strategy used

"We learn from each other continually"	*Director 7*
"More opinions is hopefully the best option"	*Film reader 44 – Advanced Practitioner*
"Don't like group arbitration and it is a confrontational, negative way to start the day"	*Film reader 12 - Radiologist*
"Radiologists frequently dismiss subtle findings that are picked up by Radiographers. The units recall rate is low and recalling these subtle areas may increase our cancer detection. We have (rarely) had interval cancers present where a radiographer had recalled but consensus was RR"	*Film reader 4 – Advanced Practitioner*
"An element of discomfort in the consensus meeting had helped reduce recall rates in the past"	*Director 9*
"Consensus meetings can sometimes become quite heated - depending on the attitude of some radiologists"	*Film reader 30 – Consultant Radiographer*

Table 28-1E Free Text Survey Comments 6.9.7 Scheduling of Arbitration/Consensus meetings

"Random depends on who is film reading"	*Film reader 15 - Radiologist*
"Ad-hoc arbitration meetings are held approximately twice per week"	*Film reader 54 - Radiologist*
"As and when required"	*Film reader 33 – Consultant Radiographer*
"Sometimes will be done between clinics"	*Film reader 5 – Advanced Practitioner*
"No set time or day"	*Director 25*
"Attempt to perform daily when staff available"	*Director 15*
"Whenever readers are available to meet"	*Film reader 52 – Consultant Radiographer*
"When we can meet up…"	*Director 18*
"During lunchtimes or any time when there are sufficient staff free"	*Director 16*
"General workload"	*Director 7*
"When clinical commitments permit"	*Director 26*
"Usually before MDT but not always possible depending on clinic workload."	*Film reader 22 - Radiologist*
"They are often held during lunchtime which is not popular as our breaks are often far less than the half-hour allocated"	*Film reader 42 – Consultant*

	Radiographer
"Usually rushed, interrupted sessions during lunch or after work"	*Film reader 52 – Consultant Radiographer*
"Extremely challenging - we hold these meetings between clinics or at the end of the day if clinics finish early'	*Director 24*
"We have set teams on days"	*Director 30*

Table 28-1F Free Text Survey Comments 6.9.8 Consensus practice

"At least one radiologist/clinician was required"	*Film reader 2 – Advanced Practitioner*
"At least two readers, neither of whom should be the original readers. Original readers may be involved if there are more readers present for discussion"	*Director 9*
"If the film reader has recalled and is part of the consensus group, they would refrain from giving an opinion on that case"	*Film reader 58 – Radiologist*

Table 28-1G Free Text Survey Comments 6.9.9. Decision making strategy at a consensus

"Recall if both agree If not, then sent to arbitration the following day"	*Film reader 29 – Radiologist*
"Majority decision, but will recall if one individual feels very strongly"	*Director 3*
"SOPS for consensus are that the recalling radiologist states level of concern 1-5 and also draws the site of concern and feature type. At consensus it is useful to know, and we do pay attention"	*Film reader 20 – Radiologist*
"Generally, all in agreement regardless of profession. Experience plays a part"	*Director 22*

Table 28-1H Free Text Survey Comments 6.9.10 Team dynamics within consensus groups

"We have an excellent team approach"	*Film reader 14 – Advanced Practitioner*
"Decision used to be made by Radiologists alone within consensus setting. With introduction of radiographer film reading, decision now ultimately heavily influenced by Consultant Radiologist/Radiographer opinion"	*Film reader 56 – Advanced Practitioner*
"Our consensus is a fair discussion where everyone's opinion is equally valid and majority opinion rules"	*Film reader 52 – Consultant Radiographer*

"Sometimes one person with a strong personality can dominate the meeting if not careful; this can cause a problem"	*Film reader 46 – Advanced Practitioner*
"Radiologist often overrules if their option differs from the rest of the consensus group"	*Film reader 57 – Advanced Practitioner*
"Despite FRQA results which show our radiographers to be as at least as good as and, in many cases, more accurate image readers, radiologists and the consultant radiographer still overrule the opinions and concerns of advanced practitioners simply because they can. Pt safety is not considered when discharging patients who have indeterminate features"	*Film reader 57 – Advanced Practitioner*
"Strength of personality of other members of the group may influence a decision to discharge by the lead, but this often becomes a professional disagreement between APs and radiologist and many APs will not contradict regardless of their level of suspicion"	*Film reader 57 – Advanced Practitioner*

Table 28-1I Free Text Survey Comments 6.10.2 Professionals currently co-ordinating/leading consensus meetings

"There is no lead of consensus; all members are equal"	*Director 23*
"No one person is taking the lead"	*Film reader 61 – Consultant Radiographer*
"Put in under CON"	*Film reader 16 – Consultant radiographer*
"Consensus decisions are entered on NBSS during the consensus meetings under 'DIS'"	*Director 23*

"Goes under Radiologist code"	*Director 25*
"Although decision is recorded as third reader arbitrator, we have a group discussion"	*Director 24*

Table 28-1J Free Text Survey Comments 6.10.3 Timeframe for Radiographer arbitration/lead of consensus

"Commenced approximately four years ago. I do not have the exact date"	*Director 5*
"More than five years ago. Though only performed by radiographers with at least one year's film reading experience, post-qualification, and approved by the clinical director"	*Film reader 25 – Advanced Practitioner*
"More than three years ago"	*Film reader 16 – Consultant Radiographer*
"More than five years ago Consultant radiographer does perform this task"	*Film reader 33 – Consultant Radiographer*

Appendix 9. KC62 4-Year Data reviewed.

Unit Name	overall recall rate 2013-2014	overall recall rate 2014-2015	overall recall rate 2015-2016	overall recall rate 2016-2017	4 yr Average overall recall rate	prevalent recall rate 2013-2104	prevalent recall rate 2014-2105	prevalent recall rate 2015-2106	prevalent recall rate 2016-2017	4 yr average prevalent recall
	4.4	5.6	5.0	4.2	4.8	9.7	10.7	11.7	9.6	10.4
	3.7	4.2	5.0	4.2	4.3	7.3	8.3	9.3	8.1	8.2
	5.2	5.8	5.2	5.2	5.3	8.9	9.9	10.9	9.1	9.7
	3.7	2.9	3.7	3.0	3.3	7.3	8.3	9.3	5.2	7.5
	3.9	4.2	3.9	3.4	3.9	8.6	9.6	10.6	7.8	9.2
	4.0	4.2	3.8	4.5	4.1	9.2	10.2	11.2	10.1	10.2
	2.7	2.4	3.8	4.2	3.3	5.0	6.0	7.0	7.8	6.5
	3.6	4.0	3.8	3.8	3.8	8.6	9.6	10.6	8.3	9.3
	4.9	3.9	4.1	5.2	4.5	9.2	10.2	11.2	10.4	10.3
	4.3	3.8	4.5	4.4	4.2	7.9	8.9	9.9	7.3	8.5
	3.4	2.7	2.5	2.5	2.8	8.6	9.6	10.6	6.1	8.7
	2.5	2.2	2.5	2.6	2.5	5.2	6.2	7.2	6.5	6.3
	4.3	4.1	4.3	4.2	4.2	7.8	8.8	9.8	8.2	8.7
	4.4	4.0	3.7	3.7	3.9	9.6	10.6	11.6	6.8	9.7
	2.5	2.3	2.5	2.7	2.5	5.3	6.3	7.3	5.9	6.2
	4.6	3.0	2.7	2.7	3.3	7.9	8.9	9.9	6.1	8.2
	5.0	5.2	5.0	5.5	5.2	8.9	9.9	10.9	10.5	10.0
	4.2	4.3	4.7	5.1	4.6	8.3	9.3	10.3	10.7	9.7
	3.7	3.9	4.5	4.9	4.3	6.8	7.8	8.8	10.8	8.5
	4.1	3.9	4.7	4.5	4.3	6.7	7.7	8.7	8.6	8.0
	3.4	3.7	4.5	4.0	3.9	6.0	7.0	8.0	6.4	6.9
	2.3	2.2	2.0	2.0	2.1	6.3	7.3	8.3	5.3	6.8
	3.4	3.1	3.2	3.0	3.2	6.1	7.1	8.1	6.7	7.0
	2.9	3.3	2.5	2.9	2.9	5.9	6.9	7.9	4.6	6.3
	2.7	2.9	2.9	3.1	2.9	7.0	8.0	9.0	7.3	7.8
	3.8	4.1	4.1	3.8	3.9	8.0	9.0	10.0	8.0	8.8
	4.7	4.5	3.9	3.1	4.1	9.1	10.1	11.1	6.0	9.0
	5.7	4.8	4.8	4.9	5.0	13.2	14.2	15.2	11.0	13.4
	2.5	2.1	2.1	2.5	2.3	5.1	6.1	7.1	4.4	5.6
	2.9	3.0	3.0	2.8	2.9	4.4	5.4	6.4	3.5	4.9
	5.4	5.2	4.6	4.5	4.9	14.6	15.6	16.6	8.8	13.9
	4.9	4.1	3.4	3.2	3.9	16.7	17.7	18.7	8.0	15.3
	2.9	2.3	2.2	2.1	2.4	5.0	6.0	7.0	4.9	5.8
	3.8	3.5	3.1	3.6	3.5	12.4	13.4	14.4	7.0	11.8
	4.5	4.3	4.1	4.9	4.5	9.3	10.3	11.3	9.5	10.1
	3.7	3.9	3.6	3.3	3.6	5.9	6.9	7.9	5.5	6.5
	3.1	3.0	3.0	3.0	3.0	6.4	7.4	8.4	7.5	7.4
	3.8	4.1	4.0	3.7	3.9	6.8	7.8	8.8	7.5	7.8
	4.8	5.0	5.0	4.3	4.8	10.9	11.9	12.9	7.9	10.9
	4.9	5.6	4.5	4.0	4.7	10.7	11.7	12.7	7.7	10.7
	3.5	3.9	4.1	4.4	4.0	6.3	7.3	8.3	8.8	7.7
	4.0	3.8	3.6	3.4	3.7	6.7	7.7	8.7	6.3	7.4
	2.1	2.3	2.1	2.4	2.2	4.4	5.4	6.4	5.9	5.5
	4.3	4.0	4.4	4.4	4.3	8.1	9.1	10.1	10.8	9.5
	4.9	3.4	4.0	4.1	4.1	11.5	12.5	13.5	8.7	11.6
	4.5	3.6	3.7	3.4	3.8	10.1	11.1	12.1	6.9	10.0
	3.0	2.8	2.6	2.5	2.7	6.0	7.0	8.0	5.0	6.5
	4.7	4.7	4.1	3.3	4.2	10.0	11.0	12.0	6.9	10.0
	4.9	4.9	4.9	4.6	4.8	9.5	10.5	11.5	9.4	10.2
	2.7	2.6	2.6	2.7	2.7	4.7	5.7	6.7	5.7	5.7
	4.0	4.2	5.2	4.6	4.5	9.1	10.1	11.1	12.2	10.6
	3.5	3.1	2.3	2.5	2.9	6.5	7.5	8.5	5.4	7.0
	3.7	3.5	3.8	4.0	3.7	7.3	8.3	9.3	9.1	8.5
	3.1	4.1	4.0	3.7	3.7	5.0	6.0	7.0	7.4	6.4
	2.4	2.4	2.6	2.5	2.5	5.0	6.0	7.0	5.0	5.7
	4.3	5.5	4.0	3.9	4.4	9.0	10.0	11.0	8.5	9.6
	4.3	3.6	3.9	3.9	3.9	7.4	8.4	9.4	6.6	7.9
	3.7	3.3	3.5	3.6	3.5	7.1	8.1	9.1	9.9	8.6
	5.0	4.9	3.9	4.4	4.5	11.2	12.2	13.2	10.8	11.9
	3.5	3.8	3.7	4.8	3.9	5.2	6.2	7.2	8.0	6.7
	3.1	2.4	2.7	3.1	2.8	6.9	7.9	8.9	7.5	7.8
	3.9	2.9	3.1	3.0	3.2	9.9	10.9	11.9	8.1	10.2
	2.5	2.0	2.4	2.5	2.3	5.2	6.2	7.2	5.8	6.1
	4.3	3.8	4.3	3.7	4.0	9.7	10.7	11.7	7.8	10.0
	5.1	5.3	4.3	4.2	4.7	11.3	12.3	13.3	7.9	11.2
	3.4	3.0	3.6	3.7	3.5	7.0	8.0	9.0	7.8	7.9
	2.1	2.2	2.5	2.2	2.2	4.5	5.5	6.5	3.8	5.1
	4.1	5.2	4.2	4.8	4.6	9.1	10.1	11.1	10.6	10.2
	2.9	3.4	3.7	3.5	3.4	5.7	6.7	7.7	6.5	6.6
	5.3	5.8	5.7	4.9	5.4	9.7	10.7	11.7	9.6	10.5
	4.7	4.3	4.7	5.2	4.7	10.4	11.4	12.4	10.5	11.2
	4.9	4.6	4.2	4.6	4.6	9.8	10.8	11.8	9.5	10.4
	3.8	4.5	4.2	3.4	4.0	7.7	8.7	9.7	6.6	8.2
	3.5	2.4	2.7	2.8	2.8	7.6	8.6	9.6	6.4	8.1
	4.6	5.2	4.9	5.9	5.2	10.2	11.2	12.2	12.8	11.6
	4.4	4.2	4.7	4.4	4.4	8.5	9.5	10.5	8.8	9.3
	3.5	3.8	3.0	2.7	3.2	6.8	7.8	8.8	7.4	7.7
	3.1	2.9	2.7	2.9	2.9	6.3	7.3	8.3	6.1	7.0
	6.1	7.5	7.0	7.0	6.9	12.1	13.1	14.1	14.7	13.5
	4.9	3.8	3.2	3.2	3.8	9.5	10.5	11.5	5.9	9.3

incident recall 2013-2014	incident recall 2014-2015	incident recall 2015-2016	incident recall 2016-2017	4 yr avearge incident recall	prevalent <15mm 2013-2014	prevalent <15mm 2014-2015	prevalent <15mm 2015-2016	prevalent <15mm 2016-2017	4 yr average prevalent <15mm
3.3	4.0	4.2	3.5	3.7	2.1	4.2	3.4	5.9	3.9
2.8	3.7	4.2	3.4	3.5	1.7	3.0	3.7	3.6	3.0
4.0	4.4	4.2	4.3	4.2	2.4	1.7	2.7	2.2	2.3
3.0	2.3	3.1	2.7	2.8	1.3	2.6	1.9	2.4	2.1
2.8	3.0	3.1	2.8	2.9	2.0	2.8	2.6	1.7	2.3
3.1	3.4	3.0	3.6	3.3	3.1	4.4	2.9	5.3	3.9
1.9	1.8	3.4	3.9	2.8	2.4	2.4	7.1	3.3	3.8
3.0	3.3	3.0	3.1	3.1	5.4	2.5	3.7	2.3	3.5
3.2	2.5	3.0	3.2	3.0	2.5	2.4	2.7	2.4	2.5
3.5	3.0	3.6	3.8	3.5	3.2	7.7	1.4	5.1	4.3
2.3	2.1	1.8	1.9	2.0	2.0	3.5	2.2	3.2	2.7
1.9	1.6	2.0	2.0	1.9	1.5	1.2	1.2	3.1	1.7
3.2	3.4	3.7	3.8	3.5	2.2	3.4	5.9	3.4	3.7
3.3	3.2	3.2	2.9	3.2	2.7	0.9	2.1	3.3	2.2
1.8	1.9	2.1	2.2	2.0	2.4	2.3	4.4	2.8	2.9
3.4	2.8	2.1	2.1	2.6	3.1	1.1	2.0	4.2	2.6
3.9	4.1	4.1	4.6	4.2	3.1	4.7	3.5	2.2	3.4
2.9	3.5	4.0	4.1	3.6	3.7	3.4	2.0	4.5	3.4
2.9	3.2	3.5	3.8	3.4	3.8	2.6	3.1	1.7	2.8
3.5	3.2	4.3	3.8	3.7	1.8	3.1	4.1	4.8	3.5
2.8	3.1	4.1	3.6	3.4	4.0	2.9	3.7	3.6	3.6
1.9	2.1	1.8	1.8	1.9	2.3	1.6	3.7	1.6	2.3
2.4	2.3	2.6	2.4	2.4	3.6	2.4	1.5	3.4	2.7
2.1	2.6	1.9	2.4	2.2	4.6	2.6	4.2	1.7	3.3
2.3	2.6	2.6	2.8	2.6	2.2	4.2	5.2	5.6	4.3
3.1	3.4	3.2	3.2	3.2	2.1	3.4	2.1	1.8	2.4
3.3	3.3	3.0	2.3	3.0	4.2	3.8	4.3	4.8	4.3
4.4	3.6	3.7	3.8	3.9	3.9	3.3	5.0	5.2	4.3
1.9	1.7	1.6	2.1	1.8	2.9	2.7	2.5	3.4	2.9
2.2	2.6	2.6	2.5	2.5	5.0	2.8	3.4	1.6	3.2
4.6	4.6	4.0	3.6	4.2	4.7	-	3.2	6.5	3.6
4.1	3.7	3.1	2.8	3.4	-	3.0	6.3	4.5	3.4
2.4	1.9	1.9	1.9	2.0	3.0	3.6	2.3	3.7	3.1
3.1	3.2	2.9	3.3	3.1	11.2	-	5.7	2.5	4.9
3.4	3.5	3.4	4.1	3.6	4.7	6.4	4.5	4.1	4.9
3.0	3.0	3.1	2.9	3.0	2.7	4.8	2.9	3.9	3.6
2.5	2.3	2.4	2.4	2.4	3.7	3.4	3.3	3.4	3.5
2.8	3.0	3.0	3.0	2.9	2.3	2.9	1.6	3.5	2.6
3.8	3.9	4.0	3.4	3.8	2.5	4.5	1.9	4.2	3.3
4.2	4.7	3.6	3.3	3.9	4.1	2.9	2.3	1.5	2.7
2.5	3.0	3.6	3.6	3.2	2.4	4.4	4.8	4.6	4.0
3.1	3.1	3.0	2.8	3.0	3.8	3.9	2.9	4.7	3.8
1.9	2.1	1.9	2.2	2.0	1.6	2.4	1.8	2.1	2.0
3.6	3.3	3.8	3.7	3.6	3.7	2.6	6.6	6.1	4.8
3.3	2.2	2.6	3.4	2.9	3.5	2.9	2.4	4.4	3.3
3.1	2.6	2.9	2.5	2.8	4.2	0.9	4.3	1.9	2.8
2.5	2.3	2.3	2.1	2.3	1.8	1.3	2.1	1.3	1.7
3.4	3.4	3.0	2.7	3.1	2.6	4.4	2.5	1.4	2.7
3.8	3.5	3.7	3.5	3.6	1.8	2.5	1.3	2.3	2.0
2.1	2.4	2.0	2.1	2.1	2.3	2.0	2.9	0.9	2.0
2.7	3.6	4.1	3.4	3.4	2.0	4.6	1.6	5.9	3.5
2.7	2.6	2.0	2.1	2.3	1.7	3.3	3.0	4.6	3.2
2.8	3.1	3.4	3.3	3.2	2.5	3.2	2.4	2.2	2.6
2.7	3.5	3.4	3.2	3.2	2.7	5.8	2.7	5.6	4.2
1.8	2.0	2.2	2.1	2.0	2.2	3.8	3.7	2.9	3.2
3.0	4.0	3.2	2.9	3.3	2.5	5.8	2.0	4.1	3.6
3.3	2.9	3.2	3.3	3.2	4.6	3.4	2.6	2.3	3.2
2.6	2.8	3.2	3.0	2.9	2.6	3.2	4.9	3.0	3.4
3.5	3.4	2.8	3.2	3.2	4.9	3.8	3.9	3.7	4.1
3.2	3.0	3.5	4.3	3.5	1.7	4.5	4.5	4.4	3.8
2.2	1.9	2.2	2.5	2.2	4.7	2.4	2.2	4.4	3.4
2.7	2.2	2.3	2.0	2.3	4.4	3.3	5.1	3.8	4.1
1.7	1.6	1.9	2.1	1.8	4.2	3.2	2.9	7.1	4.3
3.4	2.9	3.3	2.7	3.1	2.4	2.0	1.1	3.1	2.1
3.8	4.0	3.6	3.5	3.7	6.3	2.7	7.6	4.3	5.2
2.3	2.2	2.5	2.6	2.4	2.2	2.7	2.2	2.6	2.4
1.3	1.5	1.8	1.7	1.6	2.0	1.3	3.8	3.1	2.5
2.9	4.0	3.3	3.8	3.5	1.2	3.1	7.4	4.0	3.9
2.2	2.8	2.7	2.8	2.6	4.7	3.1	3.5	2.5	3.5
3.8	3.8	3.6	3.2	3.6	3.8	2.6	1.8	2.2	2.6
3.1	3.1	3.7	4.2	3.5	3.4	3.4	4.1	3.2	3.5
3.6	3.5	3.3	3.7	3.6	1.8	0.4	2.2	1.9	1.6
3.0	3.8	3.5	2.9	3.3	3.7	2.9	5.0	1.8	3.3
2.3	1.6	2.1	2.1	2.0	2.5	2.1	4.4	1.1	2.5
3.6	4.3	4.1	5.2	4.3	1.7	3.4	4.5	1.5	2.8
3.1	2.9	3.4	3.3	3.1	2.0	2.1	2.8	1.6	2.1
2.9	3.3	2.4	2.4	2.8	1.4	-	1.6	-	0.7
2.2	2.3	2.2	2.4	2.3	2.4	2.9	2.1	6.1	3.3
4.7	6.1	5.6	5.8	5.5	2.5	3.0	4.1	2.9	3.1
3.6	3.0	2.5	2.5	2.9	2.2	3.7	3.4	2.2	2.9

prevalent SDR 2013-2104	prevalent SDR 2014-2015	prevalent SDR 2015-2016	prevalent SDR 2016-2017	4 yr average prevalent SDR	Incident <15mm 2013 2014	Incident <15mm 2014- 2015	Incident <15mm 2015- 2016	Incident <15mm 2016- 2017	4 yr average incident <15mm
1.54	1.85	1.50	2.11	1.75	3.6	3.5	3.7	3.5	3.6
1.34	1.34	1.79	1.40	1.47	3.2	3.7	3.8	2.8	3.4
1.40	1.41	1.37	0.95	1.28	2.6	2.9	2.6	2.6	2.7
1.58	1.57	0.47	1.45	1.27	2.9	4.3	2.7	2.8	3.2
1.62	1.49	1.37	1.60	1.52	2.4	2.1	2.7	2.9	2.6
1.83	2.46	1.40	1.83	1.88	2.2	4.2	3.4	4.5	3.6
1.60	0.99	2.03	1.37	1.50	2.3	3.5	3.6	4.2	3.4
2.29	1.87	2.21	1.21	1.89	2.7	3.5	3.7	4.7	3.7
1.32	1.30	1.26	1.45	1.33	2.9	2.9	3.0	3.0	2.9
1.34	1.83	1.31	3.16	1.91	3.9	2.9	4.7	2.7	3.5
1.68	1.91	1.23	1.93	1.69	2.5	2.7	2.7	3.3	2.8
1.43	1.69	1.14	1.54	1.45	2.5	3.1	3.4	2.4	2.9
1.48	1.67	2.12	1.27	1.63	3.9	3.9	3.4	3.5	3.7
1.91	0.84	1.78	0.77	1.33	2.3	2.9	2.2	3.1	2.6
1.42	2.11	1.75	1.67	1.74	2.4	3.3	3.2	3.0	3.0
1.72	1.27	1.75	1.76	1.62	4.3	3.9	3.1	4.2	3.9
1.64	2.07	1.80	1.56	1.77	2.4	4.1	3.5	3.7	3.4
1.86	1.36	1.30	2.18	1.67	3.7	4.5	4.0	3.4	3.9
1.73	1.81	1.84	1.29	1.67	3.1	4.2	2.4	3.0	3.2
1.58	2.08	1.11	1.65	1.61	3.7	3.6	3.8	3.4	3.6
1.89	1.49	1.39	1.90	1.67	2.3	3.2	4.9	3.8	3.6
0.46	1.72	1.47	1.25	1.23	3.3	3.4	3.6	3.7	3.5
2.17	1.99	1.16	2.10	1.86	3.0	2.9	2.9	2.8	2.9
2.35	1.31	2.52	0.68	1.72	1.9	4.0	4.1	4.2	3.6
0.98	1.22	1.80	2.00	1.50	2.6	3.7	3.6	3.1	3.3
1.49	2.04	1.35	1.43	1.58	1.4	3.0	3.2	3.3	2.7
1.75	1.91	1.74	1.92	1.83	4.6	4.2	5.1	3.6	4.4
1.64	1.49	1.91	2.00	1.76	4.0	4.0	4.4	3.9	4.1
1.48	1.76	1.53	1.87	1.66	2.3	3.1	3.8	3.5	3.2
1.97	1.31	1.55	0.78	1.40	2.7	2.8	3.6	3.8	3.2
2.21	1.50	1.77	2.14	1.90	3.6	2.7	3.1	1.8	2.8
1.39	1.82	1.32	1.35	1.47	2.1	3.7	4.3	2.8	3.2
1.98	2.21	1.98	1.67	1.96	3.1	3.0	3.2	3.0	3.1
3.84	0.45	1.58	1.52	1.85	2.3	3.4	4.2	3.8	3.4
1.90	2.19	1.70	1.53	1.83	2.6	3.2	3.4	3.9	3.3
1.35	2.08	1.42	1.65	1.63	3.1	3.3	4.5	3.8	3.7
1.64	1.51	1.49	2.05	1.67	3.6	2.9	3.5	3.8	3.4
1.77	1.73	1.49	2.04	1.76	2.6	3.3	2.7	2.6	2.8
0.94	1.96	1.47	2.22	1.65	2.9	3.2	4.0	4.0	3.5
1.76	1.75	2.16	1.05	1.68	2.6	3.0	3.7	3.2	3.1
1.53	2.01	1.63	1.65	1.71	2.9	1.8	3.5	3.6	3.0
1.63	1.78	1.83	2.38	1.90	3.7	3.8	3.3	2.9	3.4
0.98	1.68	1.71	1.98	1.59	2.6	3.1	2.8	3.0	2.9
1.37	1.03	2.24	2.21	1.71	2.2	2.9	2.7	2.9	2.6
1.71	1.69	1.51	1.85	1.69	3.0	2.7	2.9	3.3	3.0
1.84	1.47	2.15	0.93	1.60	2.1	3.7	3.7	2.8	3.1
0.92	0.66	1.35	1.48	1.10	2.9	3.3	3.5	3.5	3.3
1.78	1.93	1.49	1.40	1.65	2.4	3.5	3.2	3.6	3.2
1.48	1.45	0.77	1.25	1.24	1.6	2.5	2.8	2.4	2.3
1.17	1.23	1.62	0.63	1.16	4.1	4.1	2.0	2.9	3.3
1.24	1.41	1.43	2.06	1.54	2.9	3.2	3.9	3.6	3.4
1.70	1.69	1.27	1.64	1.57	2.6	2.8	3.2	3.1	2.9
1.14	1.76	1.30	1.32	1.38	2.6	3.7	3.4	3.2	3.2
1.31	1.70	1.30	2.55	1.72	1.9	3.1	3.1	4.0	3.0
1.19	1.36	1.34	1.57	1.36	3.3	3.8	4.3	4.4	3.9
1.35	1.93	0.95	1.62	1.46	3.2	3.3	3.5	2.9	3.2
1.62	1.41	1.56	0.96	1.39	3.5	3.0	3.7	3.5	3.4
1.00	1.66	1.26	1.19	1.28	2.1	3.8	3.9	3.1	3.2
2.46	1.84	1.77	2.52	2.15	2.1	3.8	3.2	2.4	2.9
0.89	1.87	1.74	1.39	1.47	2.3	3.9	2.5	4.7	3.4
2.20	1.53	1.31	2.26	1.83	3.1	2.6	3.6	3.5	3.2
2.15	1.39	2.34	1.56	1.86	2.1	3.5	4.1	3.0	3.2
1.50	1.62	1.47	2.04	1.66	2.7	2.9	3.3	4.2	3.3
2.40	1.41	0.78	1.43	1.51	2.0	2.3	4.0	3.0	2.8
1.92	1.81	2.78	1.93	2.11	3.3	3.2	5.0	2.9	3.6
1.74	1.16	1.02	1.35	1.32	1.9	3.6	2.7	2.9	2.8
1.21	0.86	1.78	1.47	1.33	1.8	2.9	2.9	3.6	2.8
1.28	2.10	2.27	1.71	1.84	3.2	3.9	3.7	3.6	3.6
2.27	1.53	2.01	1.25	1.76	3.7	4.4	3.7	4.0	3.9
2.11	1.83	1.45	1.64	1.76	2.3	3.1	3.0	3.2	2.9
1.96	1.52	1.76	2.03	1.82	3.7	3.5	4.8	4.8	4.2
1.67	1.16	1.46	1.58	1.47	2.4	2.4	2.6	2.8	2.5
2.00	1.31	2.11	1.28	1.68	2.6	2.9	3.0	2.9	2.8
0.91	1.64	1.48	1.12	1.29	2.7	3.1	2.2	2.5	2.7
1.42	2.20	1.52	1.55	1.67	3.5	3.2	3.6	2.9	3.3
1.10	1.04	1.35	1.31	1.20	2.3	2.8	2.2	2.4	2.4
1.00	-	1.02	1.74	0.94	3.5	4.5	4.6	5.2	4.4
1.65	1.36	0.85	2.29	1.54	2.3	3.4	3.0	3.2	3.0
1.89	1.72	1.79	2.24	1.91	2.0	3.8	3.3	4.1	3.3
1.97	1.63	1.35	1.13	1.52	3.8	3.5	3.1	3.8	3.5

Incident SDR 2013 2014	Incident SDR 2014- 2015	Incident SDR 2015- 2016	Incident SDR 2016- 2017	4 yr average incident SDR
1.55	1.65	1.84	1.57	1.65
1.32	1.73	1.87	1.29	1.55
1.23	1.66	1.41	1.19	1.37
1.28	1.55	1.63	1.22	1.42
1.33	1.34	1.47	1.46	1.40
1.16	1.75	1.38	1.64	1.49
1.05	1.40	1.40	1.55	1.35
1.37	1.55	1.68	1.95	1.64
1.31	1.34	1.47	1.33	1.36
1.67	1.82	1.84	1.82	1.79
1.29	1.34	1.35	1.47	1.36
1.36	1.34	1.51	1.29	1.38
1.50	1.45	1.43	1.47	1.46
1.50	1.50	1.15	1.26	1.35
1.33	1.51	1.54	1.44	1.45
1.72	1.51	1.50	1.58	1.58
1.20	1.60	1.63	1.60	1.51
1.70	1.90	1.83	1.61	1.76
1.47	2.06	1.57	1.47	1.64
1.60	1.66	1.68	1.79	1.68
1.33	1.33	1.81	1.86	1.58
1.49	1.62	1.28	1.44	1.46
1.59	1.55	1.66	1.23	1.51
1.25	1.92	1.82	1.72	1.68
1.30	1.44	1.56	1.42	1.43
1.08	1.65	1.67	1.70	1.52
1.89	1.68	2.02	1.46	1.76
1.62	1.66	1.86	1.69	1.71
1.20	1.46	1.44	1.57	1.42
1.08	1.32	1.41	1.51	1.33
1.37	1.66	1.68	1.29	1.50
1.49	1.70	1.57	1.12	1.47
1.51	1.62	1.66	1.40	1.55
1.52	1.55	1.82	1.49	1.60
1.30	1.37	1.32	1.62	1.40
1.40	1.43	1.58	1.50	1.48
1.51	1.34	1.54	1.49	1.47
1.40	1.54	1.52	1.40	1.46
1.41	1.56	1.75	1.72	1.61
1.20	1.50	1.73	1.54	1.49
1.43	1.12	1.66	1.67	1.47
1.81	1.65	1.59	1.38	1.61
1.26	1.48	1.23	1.42	1.35
0.99	1.39	1.24	1.23	1.21
1.46	1.19	1.12	1.40	1.29
1.12	1.42	1.30	1.10	1.23
1.25	1.20	1.48	1.51	1.36
1.27	1.57	1.53	1.76	1.53
1.12	1.43	1.48	1.36	1.35
1.34	1.79	1.07	1.24	1.36
1.19	1.14	1.57	1.48	1.35
1.39	1.38	1.45	1.42	1.41
1.20	1.60	1.63	1.50	1.48
1.22	1.29	1.52	1.72	1.44
1.19	1.62	1.52	1.57	1.47
1.66	1.61	1.35	1.22	1.46
1.41	1.35	1.50	1.53	1.45
1.37	1.48	1.65	1.27	1.44
1.48	1.88	1.52	1.31	1.55
1.00	1.57	1.04	1.95	1.39
1.39	1.27	1.51	1.80	1.49
1.20	1.58	1.74	1.47	1.50
1.27	1.08	1.36	1.78	1.37
1.39	1.32	1.59	1.49	1.45
1.41	1.25	1.74	1.31	1.43
1.09	1.52	1.20	1.32	1.28
0.87	1.29	1.31	1.53	1.25
1.63	1.66	1.56	1.37	1.56
1.60	1.61	1.50	1.74	1.61
1.37	1.63	1.50	1.48	1.50
1.43	1.60	1.83	2.04	1.72
1.41	1.40	1.50	1.55	1.47
1.54	1.64	1.60	1.64	1.60
1.51	1.39	1.40	1.41	1.43
1.50	1.36	1.63	1.34	1.46
1.29	1.51	1.25	1.34	1.35
1.45	1.63	1.55	1.62	1.56
1.33	1.43	1.69	1.38	1.45
1.42	1.88	1.68	1.87	1.71
1.60	1.45	1.54	1.48	1.52

Appendix 10. Interview Topic Guide

Interview topics	Questions
Reporting practice/arbitration practice history.	Non-blinded reading/arbitration and effects on decision making. Effect of reading type on the number of arbitration cases. Team dynamics within consensus groups
Receptiveness to change	Opinions on fully blind reading and a paperless system – benefits/disadvantages
Defining an experienced reader/arbitrator	What would you class as an experienced reader? What are the criteria? How do you select an arbitrator?
PHE arbitration guidance	What is your opinion on the content of the guidance?
Recommendations for improvement of the guidance	Is there anything you think would improve the guidance?
Quantitative guidance for new arbitrators	How would you define quantitative guidelines for new arbitrators?
Implementation of the guidance.	Effect of the guidance Barriers/facilitators to implementation
Centralisation of arbitration services	What are your thoughts on consolidating expertise for arbitration with cases sent to a centralised service or external arbitrator?

Appendix 11. Participant Informed Consent Form

Title of Study: A study to explore arbitration and consensus processes within Breast Screening Units in England. Researcher: Lisa Hackney

	Please confirm agreement to the statements by putting your initials in the box below
I confirm that I have read and understood the Participant Information Sheet for the above study and have had the opportunity to consider the information, ask questions and have had these answered satisfactorily.	
I understand that my participation is voluntary and that I am free to withdraw consent during the interview, without giving a reason.	
I understand that all the information I provide will be treated as confidential, will be fully anonymised and stored securely.	
I agree to an audio recording of the interview to allow for anonymised transcription.	
I agree to take part in this study	
Participant Signature …… Date	
Name of Participant	
Researcher Signature …….. Date	
Name of Researcher	

Appendix 12. A table of themes, sub-themes and codes

Theme	Sub-theme	Codes	Description
Organisational factors	Organisational variance and historical, cultural elements	Historic practice/local culture	Organisational or professional culture defining departmental processes, policy, norms, tasks.
		Lack of standardisation	Evolution of practice -double reading practice, blind vs non-blind reading, double radiographer reporting, third reader arbitrator or consensus group review. Working styles presently occurring with the unit.
		Non-blind reading	Perceptions and rationale of non-blind reading, the justification for current practice.
	Conformity of practice	Following on	Conformity of readers, the act of changing their decision, the objectivity of non-blinded reading/arbitration and effect on FRQA data.
		Potential for error	The potential for perception errors, lowered attention/distractions, a duty of candour.
		Psychological factors impacting on clinical decision making and performance	Bias, attitudes, personality, cognitive factors, and behaviour. Outcome on individual and unit performance statistics.
		In-house training	Training in-house, diversity (or lack) of reader profiles.
	Silo working and the concept of centralisation	Split site working	Individuals or teams are working in co-located settings, e.g. working in different geographical locations or sites.
		Staffing levels and resources	Resources for film reading and arbitration, workload pressures.

		Perception of external arbitration	Opinions on external arbitration – positive and negative perceptions, achieving NHSBSP targets, electronic infrastructure to support external arbitration
		Silo working	Screening services operating as separate units, lack of exposure to other units reading and arbitration practices.
		Collaborative working and transferable knowledge	Partnership working with other units/trusts. Improving standards through sharing of film reading and/or arbitration, learning from external practices.
Technology	Lack of sophistication of technology to support: • Blind reading and a paperless system • Audit	Compatibility of a fully electronic system	The degree to how a fully electronic system fits with existing workflows, negative and positive comments. The level of compatibility of an electronic system with organisational values and work processes. Blind reading and paperwork (did not) need to be adapted as evidence of compatibility or lack of compatibility.
		Paper system as a failsafe	Perceived risks of a fully electronic system, benefits of maintaining a paper system.
		Onerous administrative tasks	Reporters are completing paperwork, correlating numbers in film batches, separating recall cases from normal cases and technical recalls.
		Tension for Change	The degree to which participants perceive NBSS requires a change. A strong need that the current system is untenable, e.g., statements that the IT infrastructure changes are necessary, clinical trials are requiring blind reading.
		Lack of standardised recording of personnel at consensus	Variances in the recording of personnel contributing to the consensus decision.

		Feedback and audit	The degree to which reader performance is communicated and fed back to staff.
		Reflection and evaluation	Staff reflecting and evaluating on their performance and changing practice accordingly
	Challenges and prospects of Artificial Intelligence	Double reading as a safety net	Double vs single reading. Cancers detected by one reader only, litigation of missed cancers.
		Challenges/ barriers to implementing AI	Regulatory clearance, access to data, availability of AI, limitations of AI, trust and integration with existing systems
		Perceived Benefits of AI	Improved efficiency and effectiveness, and work processes. Technology that is compatible with national goals. Recall decision thresholds, personalisation of AI, learning from AI.
Clinician Factors	A meaningful measure of performance	Professional role	The professional role of the reporter or arbitrator.
		Defining an experienced reader	Factors relevant to the experience of the reporter or arbitrator, years of reading, the number of films read.
		Performance metrics and competence	Film reader performance monitoring and metrics to define competence, e.g. sensitivity/specificity/PPV/recall rates.
	Difficulties in defining quantitative guidelines for arbitration/selecting individuals	Third reader statistics	Staff knowledge and views on third reader statistics.
		Measurable boundaries for the task	The complexity of defining sensitivity and specificity of third reads, reflected by them being a sub-set of cases (small numbers of cases). Suggestions for measuring arbitration outcomes
		Pairing of readers	The degree to which reader combination impacts on recalled cases and subsequent assessment numbers.
		Variance in recall rates	Individual reader variation in recall rates- high or low.

		Selecting arbitrators	Performance metrics that may be useful in selecting individuals as third reader arbitrators.
	Radiographer self-efficacy	Confidence in undertaking the task	Low or high confidence. Individuals believe in their capabilities to undertake third reader arbitration (lead consensus), and the support required.
		Cautious behaviour/risk aversion	Over-recalling rather than the risk of dismissing cancer.
	Outcome expectancy and radiographer training	Training	Professional development training. Protocol-based practice, learning style. New or improved existing skills.
		Outcome	A projected outcome, i.e. using the status of current performance to reflect on future outcomes.
		FRQA data management	Managing and using FRQA data to enable individuals or teams to gain an understanding of their outcomes and improve future performance.
		Characteristics of an individual	Personal traits such as tolerance of ambiguity, motivation, competence, decisiveness.
	Decision-making skills	Continuous learning	Learning from reviewing the outcomes of cases. Act upon the learned outcomes.
		Behaviour change	Decision-making skills improving or declining (based on the known outcome of cases).
		Autonomous decision makers	The aptitude of staff to make independent decisions.
Teamwork factors	Collaborative working	Learning environment and educational opportunity	Adequate time and space for reflective thinking and evaluation. Collaboration for formal and informal learning – learn from each other, increased knowledge via consensus discussion, learning from missed cases to improve future outcomes. The degree to which consensus exhibit a "learning climate." Elements of group norms.
		Supportive structure	A climate in which consensus leads/ third reader arbitrators express their fallibility, the value of team members' input; team members feel valued in the decision-making process. An environment of trust.

	Team dynamics and collegial conflict	Non-supportive structure	A climate in which individuals do not feel valued or have the opportunity to voice their opinions. Opinions are constantly over-ridden.
		Relationships	Group dynamics – personal relationships within the team to include mutual respect, interpersonal cohesion. Dynamics affected by power relations within the group- positive or negative.
		Communication	Communication and the quality of communication between staff, poor communication, communication problems, good communication
		Group processes	Social and psychological interactions affecting behaviour, team decision making (rational and non-rational) and effectiveness. Hierarchical relationships allowed (or not) to constrain collaborative working.
		Confrontation/ conflict	Disagreements, arguments, confrontations or conflicts with colleagues
		Group composition	Factors which influence consensus practice, positive or negative - Group size, experience, professional roles, dominant personalities, groupthink.
	Accountability	Self-assurance	Confidence or lack of confidence in decision making at film reading. Peer support to confirm/negate the decision to recall.
		Responsibility	Diffusion of responsibility for the final decision on cases – positive or negative comments. Individuals (not) contributing to the same degree as if they were undertaking third reader duties
		Legislative	Fear of missing cancer, incorrect decision making.
PHE Guidance Factors	**Guideline factors**	Knowledge about the guidance	Individuals' familiarity with the criteria and principles.
	Evidence Strength &	Quality of the evidence	Staff perceptions of the quality and validity of evidence supporting the guidance. Awareness of the strength and quality of evidence, as well as the absence of evidence or a desire for different

	Quality		types of evidence.
	Clarity of the guidance	Clearness of the statements	Staff perceptions on the (lack of) confusion regarding the guidance criteria
	Individual professional factors		
	Lack of agreement	Attitude related statements	Staff attitudes toward the guidance positive or negative, role ambiguity.
	Inertia of practice	Priority of the guidance	Individuals' perception of the importance of the guidance within their organisation.
		Organisational culture	The willingness to change practice
	Appropriateness	Relevance of the guidance	The degree to which the guidance is appropriate to meet local needs.
	Implementation climate/capacity for change	Relative Advantage of delegation	Radiographer arbitration/consensus lead improved existing practice.
		Putting the guidance into practice	The capacity for change, support within the organisation, the general level of receptivity to implementing the guidance
		Back-up behaviour	Resorting to radiographer arbitration in times of high demand to ensure performance (screen to RR) is not negatively affected

Appendix 13. Table 47-1A Interviewee Quotes

Sub-Theme 1.1: Organisational Variance and Historic Cultural Elements

"some of the approaches…have really worried me because there was one unit, it doesn't do it anymore, that used to deliberately not write their real opinion as second reader in order to ensure it got to the arbitration pile automatically. So, if the first person had recalled, because they wanted all cancers to be discussed, the rule was the second one would always RR. …. it's very bizarre isn't it"	*Radiologist 3*
"I don't know whether it's just habit and because we haven't been told that we have to read blinded. It isn't a quality that we have to fulfil"	*Consultant Radiographer 4*
"It's just the way we've done it and nobody's, nobody in authority has made a decision to change it"	*Advanced Practitioner 3*
"It's the only way we've ever reported"	*Director 3*
"I suppose it's because that's the way NBSS was set up for us and nobody's ever suggested that we would read blinded. I think it was suggested once when they were doing that reverse trial where they turn the direction around of the way you read, but we still didn't do it blinded then"	*Advanced Practitioner 1*
"I think if we go back many years ago when we first started consensus, we did have two very new consultants and it was a high recall rate, so I think we did decide to do it that way round and we just stuck with it really. We've now got a different radiology team in and um, we've just really stuck with it"	*Advanced Practitioner 7*
"There's never really been a need, they would be quite happy for me to read against one of the film reading radiographers, but there's never been a need, so we don't do it….so we are quite happy just going along the way that we're doing. I don't think they would allow the other two radiographers to do it (read together), they are both advanced practitioners and very experienced they have been film reading for like eight years and six years"	*Consultant Radiographer 3*

"I think it does influence your decision, but in a way, it can be positive, especially if it's a new reader because there's sort of a learning skill with that. Because they have seen what an experienced reader has called and what they haven't called back"	*Director 3*
"I think, to some extent, it does influence decision making. I think that that is a learning, that the benefit of learning from that because in particular, I know that the less experienced readers will take into account what somebody they respect opinion is before they necessarily come to a conclusion"	*Director 1*
"If you didn't see something and you see that the first person did, you can learn from that on the spot"	*Consultant Radiographer 4*
"I think it's useful in many respects to be honest because film readings not a competition in terms of you know, I missed more than you or I've found more than you sort of thing"	*Locum Radiologist 1*
"If you know what the first reader has said if they have recalled, and you choose not to recall, you know that you need to put that case out for arbitration or consensus, if you don't know what they've said then you might put it in the wrong pile, you might put it back to routine recall. So, it relies on having an office that is good at sorting out what the readers have said and never making a mistake with that, for the right results to go ahead as normal if you are reading truly blinded"	*Radiologist 3*
"I guess the QA visitor people must notice how we do quite different things even those there's guidance, but I guess arbitration is just one of those things that is done very differently in lots of different places. You think how can it all be different when we basically doing the same thing or you think we would but there's very big differences between places – odd"	*Radiologist 2*

Table 47-1B Interviewee Quotes from Sub-Theme 1.2: Conformity of practice

"Personally, I think it has quite a big impact. I think it puts you under quite a lot of pressure to make the call to recall it, and it has the potential for dominant people to influence recall rates one way or the other"	*Consultant Radiographer 2*
"I'm sure that it does affect it (decision making), so I think it increases your recall rate and not necessarily in a good way"	*Radiologist 2*
"If you're the first reader then everybody else sees what you've done so they can be guided by that"	*Advanced Practitioner 2*
"We used to actually mark-up, we'd ring something, we'd circle something to show what it was we were looking at"	*Advanced Practitioner 1*
"So things that I probably wouldn't have even perceived has suddenly got big arrows on and I think that you then you start to see them in a different way, so things that I wouldn't even notice, it's then like oh yeah might that be something"	*Consultant Radiographer 2*
"I think it can affect the second reader if they're not blinded erm when you look at my statistics for cancer detection rates actually I have a lower cancer detection rate when I am second reading which sort of suggests that I am influenced by the first reader and may dismiss cases that I would otherwise have called back if I had been the first reader'	*Director 6*
"I'm much more likely to recall a case that the first reader has recalled than I am to recall a case without an opinion. So, my first reading recall rate is lower than my second reading recall rate"	*Radiologist 3*
"So yes, there are times when I've gone oh gosh yes, I can see it now, and I agree yes that's a cancer that I would have missed had I been the first reader"	*Advanced Practitioner 3*
"If you don't recall it, and then you realise at the end oh I have disagreed with somebody unknowingly you go back and look at it, and you might think oh right I just didn't see that so I think there would be some, where you just genuinely haven't perceived what they have seen"	*Director 2*

"To be honest if I think oh god there's you know I had my eyes shut during one of them or whatever it's silly to put it through to a 3rd reader if I totally agree with the first reader but we don't have a unit policy you know to say we will have it blinded"	*Radiologist 1*
"When you are looking at things at assessments… I would say that I do not believe that both people saw the abnormality and I think it's the fact that they have been directed to this area that I just don't believe that two people would perceive them as an abnormality if they did it blind"	*Consultant Radiographer 2*
"At the moment if there's doubt for a, not an experienced, a junior reader sometimes the fall-back is to go and see if somebody else has recalled it and if they haven't then follow suit and not to recall it"	*Director 4*
"The younger ones the one's that perhaps haven't been qualified so long might be inclined to just go along with what somebody else has said, but then you could say that's just perhaps they are being influenced by somebody they feel has more experience than them"	*Advanced Practitioner 1*
"I know the other radiographer reporters they always say that it's so much quicker to second read because if you're on the fence and you see that someone else has put it, I think you would, you naturally would just put your name on it as well"	*Consultant Radiographer 3*
"I think when you do the meetings as a group, you can be swayed a little bit because almost like you would be if you're non blinded second reading. So as soon as someone says oh that's there, you go oh yeah, yeah. I think that there are disadvantages in that"	*Consultant Radiographer 4*
"Sometimes they are influenced by the experience of that reader, so they don't stick to their sort of original decision making, and they change their mind to coincide with a more experienced reader"	*Director 3*
"I do think there is a tendency with some other readers to perhaps take a slightly lazy route. If they see that somebody else has abnormaled something, they just go straight to the proforma ….and think oh yeah, I'll agree with that and then they don't bother really reading the images at all they just agree with person number one. Which you know it's not very good really because of course, we know for a fact that sometimes you can be distracted from a second abnormality or something on the other side by concentrating too much on the one thing can't you?"	*Advanced Practitioner 1*

"when you're a second reader, I think sometimes people can be lazy, and assume that the first person has read them"	*Advanced Practitioner 3*
"I was speaking to the consultant radiographers, and what they told me, one of the things that influences whether they are going to recall somebody or not is dependent upon who are they going to read with which I found really interesting because I didn't even consider that. So, because we all have set days for film reading what they said is if they knew they were going to read with a certain person they were more likely to recall things that they might dismiss. So instead of having that as a case where they didn't recall it, they will recall it pre-empting what the reader is going to do as a second reader which I hadn't even considered"	*Director 3*
"I wonder maybe that's happening with the newer sort of consultants that are starting maybe they are being influenced in this way as well. And maybe over time because they do assessment clinics, they then have the confidence to make their own decisions"	*Director 3*
"I noticed that some would recall all the cases and others would make the decision. So I went round and asked each of the individual radiologists that if you were put in a situation of having to recall, a third opinion, if you disagreed you wouldn't recall it, and half of the team told me they would, and the other half said they wouldn't because they would base it on the readers that recalled. So, when they said that they wouldn't stick by what they thought, they weren't offered arbitration"	*Director 3*
"but on times arbitration you would then probably consider who's recalled it, you know your own film readers"	*Advanced Practitioner 2*
"I have worked with colleagues with a very high recall rate and what tends to happen then, I think you can run the risk of missing cancers because you think oh this is so and so, so they have over-recalled this so you know you are more likely to sort of kick it back when you shouldn't be"	*Locum Radiologist 1*
"I think it's good practice" (blinded).	*Breast Clinician 1*

"I wouldn't want it completely blinded I don't think"	*Consultant Radiographer 1*
"I think ideally we should read fully blinded"	*Consultant Radiographer 2*
"I've just been looking at BSIS for (area removed) I do know that when you look at the profiles, some of them look like fields of sheep they've got individuals all over the place and the others of the units look like shoals of fish where they're all closely packed in one area and interestingly the shoals of fish if you look at the unit data, the overall outcomes don't do as well as the fields of sheep"	*Director 1*
"If it was truly blinded, I think it would increase the arbitration rates"	*Director 4*
"I am convinced it would because if ermm if you can't see what the first reader has said"	*Radiologist 3*
"Yes, I think it would slightly because we're all going to miss something aren't we, I think it would increase arbitration numbers a bit yeah"	*Locum Radiologist 1*
"I think it would make them increase"	*Radiologist 2*
"Because we read sort of blind read, so sometimes we get annoyed that they only include first reads cause as far as we are concerned, we would like it to include all our reads because otherwise we only get stats from half our reading numbers really."	*Director 5*
(Blind reading) "I think that people who maybe do better as second readers will do worse"	*Director 3*

Table 47-1C Interviewee Quotes from Sub-Theme 1.3 Silo working and centralisation of services

"We are teetering on the edge we do appreciate that"	*Advanced Practitioner 2*
"Golly, yes. Well I mean I had not thought of sending stuff out, but I appreciate there is a real manpower issue"	*Breast Clinician 1*
"So, to actually send out through another screening service, if NBSS can be that clever and I think that would probably, everyone would put their hand up and say, oh thank God for that"	*Advanced Practitioner 2*
"You could argue that small units would be the ones to offer outside arbitration to because otherwise do they do everything by consensus, or do they take a cautious approach and if one reader calls it gets called back"	*Breast Clinician 1*
"And so you're not in your own little unit bubble, and it does get quite insular doesn't it because you have a few dominant senior people who are training the newer people coming in and people sort of converge to a fairly similar reading pattern I think and having some exposure to other units and their reading styles is probably really important"	*Director 5*
"I think that has to be the case. I don't see how small units can manage otherwise and a single practice radiology or even dual practice is not uncommon, and then if you start to cut corners when people are away, there are going to be errors. So yes, if there was some sort of central system where people could get films read or arbitrated that maybe we could all dip into that would be a good idea but a huge sort of IT nightmare I would imagine"	*Director 4*
"Some radiologists might vent well I wouldn't have called that back, you know, so ermm rather than criticising each other we've just got to get on with it isn't it"	*Director 3*
"because we're quite insular apart from the QA visit"	*Radiologist 2*

"I think the way it would work is very close ties between trusts, it being of mutual benefit so, if you had I would actually say it was best if you link up small units with each other so it's as much in their interest to do the reviews as they will then benefit from others reviewing theirs, that kind of arrangement"	*Radiologist 1*
"I think small units you are going to end up with a unit that has little, a particular flavour and you know if you've got two superb radiologists you're going to have an absolutely brilliant because you're not going to have anybody else dragging them down, but on the other hand if you've got two mediocre radiologists you're never going to achieve greatness. So, you know you take four or five small units, and you pool all the film reading"	*Director 1*
"I think I would find it quite difficult if I hadn't actually have been at the consensus meeting and, if it was outsourced then you would need very very clear guidance on exactly what it is they actually want you to look at"	*Consultant Radiographer 3*
"As long as they've got the 5,000 film reads in their pocket and everything"	*Consultant Radiographer 4*
"It would be nice to be able to send images or batches of images to other places, so you know if you have lots of people to read it would be nice to be able to send them around the country and just do them as batches"	*Radiologist 2*
"We are sort of coping here, but we can see that it would be nice to have a system where you could, support other units say in screen reading because the IT is there, you know NBSS is national"	*Locum Radiologist 1*
"I don't know, I mean who's to say that necessarily, they would, why would we do that? in a suggestion that they could do a better job at it, than the people locally"	*Advanced Practitioner 1*
"I think it might be useful for our department, but I think people would find it difficult to accept somebody else's view and not to be able to explain it out at them"	*Radiologist 1*

"I do feel strongly about this I think it's much easier to deal with a case within the same unit than somebody else outside of the unit arbitrating. A) they might not arbitrate what everybody else is, and then you've got then another problem. You know they pick up something else, or they are concerned or whatever but also it's much easier to feel confident about something that's you can follow the pathway, I suppose you could argue even if you farm out you could still follow the pathway. There is just something slightly uncomfortable it's a bit like trying to do somebody else's assessment from a different unit, but I suppose with arbitration you could argue that that's already had one read within the unit. Perhaps it wouldn't have the same impact, yes yes I've just not thought about it"	*Breast Clinician 1*
"I would not split the arbitration from the film reading. I think it's really important that you keep those in the same pool, and you want to maintain communication within that pool"	*Director 1*
"If you're then sending images off to ermm to a consensus arbitration decision and you know, everything's target driven isn't it, with NBSS updates. You have to meet this, that and everything else. So, I don't know; it's just another step in the process, isn't it?"	*Consultant Radiographer 4*
"I think I think it would; it's going to hold up the patient pathway a bit isn't it if you do that. We are already struggling because we are threatening to breach on a regular basis just trying to keep up with our reading but then if you have to send stuff away for arbitration it would be waiting even longer wouldn't they"	*Advanced Practitioner 1*
"I suppose you would have secure email or whatever, or a separate log into NBSS where you can access these cases"	*Consultant Radiographer 4*

Table 47-2A Interviewee Quotes from Sub-Theme 2.1 Lack of sophistication of the current technology

"You know maybe we're all clinging onto bits of paper as a sort of comfort blanket but ermm So somebody will have completed the number of films to be read, and then we fill in whether we have requested previous images, whether we have done if there are any TRs, and then the recalls and then the arbitrations and you know and do a bit of maths and then we also fill in a list with the names of the patients in each category and their numbers you know their screening numbers so that everybody can be certain what is happening with the patient's that haven't just gone to routine recall"	*Director 2*
"I have the pieces of paper in front of me and, and I try not to look at what the answer at what somebody else has given. I think that's the way it should be; I'd rather not see what somebody else thinks because it's my reading that I should be doing"	*Radiologist 2*
"I try and put my arm over what might be written and look at just the patient demographics from an identification point, to make sure I'm reading the right ones, and then look at them, make my decision, and then look to see whether someone's called it"	*Consultant Radiographer 1*
"at a recent Q&A visit they were saying there's all those mistakes where the readers don't sign their names, and you are thinking why are you nagging the readers to sign their names when they don't want to do anything further with this case…. the audit person was saying ooh and I have to keep going back to them because they haven't signed the paperwork and the radiologist saying it's a complete nightmare I have to sign the paperwork and you know, I know I get it wrong"	*Director 1*
"Yes, I could well see that, I am all for paperless (laughter)"	*Breast Clinician 1*

"I have to say if I was designing NBSS and I've thought long and hard about it I don't think you should have paperwork in front of you when reading at all"	*Director 1*
"Well that's how NBSS ought to be, isn't it? True paperless is where we need to be, isn't it really?	*Radiologist 3*
"I think it would be advantageous. Yep that's the way we're going isn't it you know paperless, that's the way digital health records are going"	*Director 4*
"Personally, the paperless sort of thing, I think it's safer to have all your information stored in one place. I'm not that confident with bits of paper, duplicating work so that you've got information on three or four different systems, bits of paper, I think it's how mistakes happen"	*Consultant Radiographer 2*
"No, no, I think it needs both. Because ermm checking clinic sometimes we've had radiologists have written the wrong paperwork and the screen and vice versa they've written on the paperwork the wrong result against NBSS. You won't get me to give up paper (laughter) we've gone partially paperless, and I hate it paper records are brilliant"	*Advanced Practitioner 3*
"It would be very nice being completely electronic rather than having to resort back to paper copies and stuff, it seems ridiculous in this day and age really doesn't it"	*Director 5*
"I*f you're sat there with a whole batch and you're quickly trying to get through them that certainly is going to add time on them, on to it compared to a quick cross, calc and an RC"*	*Breast Clinician 1*
"I don't see a massive problem with it (centralisation) other than the fact that logistically it might be a bit tricky"	*Consultant Radiographer 4*

"I think it could be such a powerful tool because the radiographers can indicate a problem. Which then the readers can say, yes, I've seen it because there's no guarantee at the moment that the readers have actually made any effect on the alerts that's put on"	*Advanced Practitioner 2*
"I must admit I do like having my pictures and being able to draw what I want on them rather than what the computer will allow you to. Like will there be a facility to put little dots when it's microcalcification or, we have a particular way that we draw a distortion and things like that, but ermm I think we all like our bits of paper. Probably stuck in the dark ages (laughter)"	*Consultant Radiographer 3*
"I almost know what they mean when they write things, and I think my worry would be that an electronic form I'd be thinking you know they've said it's there, I'm wondering if it's that, you know what I mean for the subtle things I think I would worry it would add to my confusion doing the assessment clinic. I mean but, in a way, it would be good"	*Radiologist 1*
"because you can't rely 100% on computer entry. You know if you make a mistake and click one of the drop-down boxes incorrectly then potentially you could have a major error"	*Locum Radiologist 1*
"So, it doesn't matter if they've got 3 years or 10 years' experience if they won't change. If by 3 years your film reading rate is still high, then even though you've had three years of experience which should have brought your film reading rate down and it hasn't then that sort of implies to me that they are not learning"	*Director 3*
"But it's all very time consuming to go back and look at all your individual disagreements other than the cancers, and again it comes back to the limited time that we all have"	*Director 2*
"We have a good collaborative interval meeting quarterly, and we all come together we review the cases and that's really good, so if I had a bit more time I think it would be nice to have a similar thing for the 3rd read cancers and kind of review them all together, that would be my ideal thing"	*Radiologist 1*

"This is probably the thing that we lack more so on than anything, is the auditing of what we actually do"	*Advanced Practitioner 2*
"I am about to try and audit all of our arbitrations to have a look at things that presented as interval cancers, but there's just finding the time to do it"	*Consultant Radiographer 2*
"We currently complete a paper record for each person anyway erm, which is a little bit of overkill"	*Locum Radiologist 1*
"My unit director tells me that she doesn't like the idea of going paperless because if you make a mistake on the computer then the mistake is the only mistake and there's no paper to disagree, and you'll never fix it"	*Radiologist 3*

Table 47-3A Interviewee Quotes from Sub-Theme 3.1 A Meaningful Measure of Performance

"Well I've done aptitude tests of film readers every time we've employed film reading radiographers, and I know from our long-term stats that those aptitude tests are exact or correct and they show me that registrars are no more able to read negatively than radiographers or anyone else. They don't have a particular gift so we should be judging people not on their title but on what they can do"	*Radiologist 3*
"I don't think it would be necessarily years or numbers of films read. Because there are people you might class as experienced because they've been doing it for years and years and years and they still miss the most obvious cancers"	*Advanced Practitioner 1*
"because experience and being good is two different things, isn't it? you could be film reading for five years, but you could be, have a 15% recall rate and a tiny cancer detection rate"	*Consultant Radiographer 4*
" I don't think experience in years makes a big difference and in fact I know people sort of can get worse later on when they you know, it all depends on family circumstances if you're having a rubbish time then reading obviously gets quite poor and so I don't know whether years in the job is really a useful thing it's your performance"	*Radiologist 2*
"One of our film readers here she's been brilliant right from the word go from the time that she was a student just learning it and I probably consider her opinion above some of the, a couple of the radiologists because she is just so good at it"	*Consultant Radiographer 3*
"Well I think there's research to show that experience in mammogram reading doesn't make you, it doesn't make you better, so you're either good at mammograms, or you're not, and whether you've had five years or fifteen years I don't think it makes a big difference…you're either going to be decent or not decent"	*Radiologist 2*

"I think the best thing would be for the best reader in your unit to do the arbitrations and I guess there is, you could argue what makes the best reader - but your highest detection rate with your lowest recall rate"	*Radiologist 2*
"we would say that they should and obviously during those 3 years they should have been reading an adequate number of images in the screening programme you know the minimum number of 4000 and five thousand in total and also we would say they should have recall rates that are below the minimum to count as an experienced film reader"	*Director 2*
"lot of it is down to individuals and their own perception because someone can be absolutely, you know, has 10 years' breast screening, you know, experience, but still with a very high recall rate"	*Advanced Practitioner 2*
"I think if you're not learning, if by 3 years your film reading rate is still high then even though you've had three years of experience which should have brought your film Reading rate down and it hasn't then that sort of implies to me that they are not learning, so it doesn't matter if they've got 3 years or 10 years' experience they won't change. So, I think for me experience is the number of years but also looking at the FRQA to show evidence that their recall rate is reasonable"	*Director 3*
"There's two things I will consider here certainly the number of years that you've been doing it and that's essentially the same as the volume of mammograms that you've reported. But there's also an element of confidence as well because erm, I don't know if that's quite the same as, I mean you can have an experienced reader that isn't actually particularly confident, lacks a bit of self-confidence, lacks a bit of what's the word erm they can be easily influenced perhaps, or more easily influenced perhaps than others. It is quite a long process I think becoming an experienced radiologist, breast screening radiologist and it's at least several years. I would have thought five years would probably be the minimum to be calling yourself an experienced film reader"	*Director 6*

"I think it would be hard to put figures on it wouldn't it, FRQA wise as in a certain recall rate or cancer detection rate that you had to meet and there is such variation isn't there from year to year and analysing that data"	*Director 5*
"5 years and reading the required number of films, undertaking assessments and PERFORMS"	*Director 7*
"I would say you probably need 5 years of pretty continuous reading with no major career breaks to be classed as an experienced reader to get that breadth of experience…. somebody who has good performance data, somebody who is performing well themselves, because you could have an experienced person who wasn't performing well"	*Director 4*
"I don't think you could class yourself as an experienced screen reader with a minimum of less than 5 years reading experience"	*Radiologist 1*
"For the experience it's almost, it isn't really a time thing, more a confidence in their own ability and to be able to make decisions that they are willing to stand by"	*Consultant Radiographer 2*
"I don't think it's determined on either the number of cases you've read or the time that you've been doing it, to be honest… some people have an ability to form an opinion if you like"	*Locum Radiologist 1*

Table 47-3B Interviewee Quotes from Sub-Theme 3.2 Difficulties in defining quantitative guidelines for arbitration/selecting individuals

"I don't think there's even a consensus view in the country about what you want an arbitrator to do. I know it sounds a bit silly, but I think the truth is exactly how sensitive and specific they should be hasn't been defined anywhere"	*Radiologist 3*
"I think it should be taken with your performance and what your sensitivity, specificity is and things because again, you could be film reading for five years, but you could be, have a 15 percent TR rate and a tiny cancer detection rate. So that doesn't necessarily, and then if you are arbitrating, you'd be calling everybody back and nothing would, you wouldn't actually increase your cancer detection rate. So, I think, I haven't got an opinion of what those rates should be specifically"	*Consultant Radiographer 4*
"I think the stats across the (name removed) when I looked at them the range of cases that were brought back varied from 25% of those sent to arbitration to over 70% of the same client. So, you're looking at that thinking well that sounds like there's a very big difference in the effect of arbitration in one unit from another. I think it's somewhere around 15% of the cancers a bit less than that, that end up in that arbitration pile that we eventually find and so I think it matters"	*Radiologist 3*
"You and I might argue that across the nation and given the number of cancers in the arbitration pile it is actually a question that matters and something that to some extent we should really be turning our attention to. So, I care about it, but I don't have all the answers"	*Radiologist 3*
"And I got my particularly good reader, and I just said I just want you to do all the arbitration please and the recall rate got cut by about 40% almost overnight. I think that reset the unit recall rate"	*Director 1*
"See what they actually do in practice and then that might tell you who you would really like to be arbitrating for you and then after they've got their arbitration licence maybe you check them every few years and let them do all the arbitrations"	*Radiologist 3*
"I don't know because I was wondering whether you should have a number that you do. If you've got to read 5,000 films a year, should you do so many arbitration cases and should you have some separate audit of that"	*Consultant Radiographer 4*

Table 47-3C Interviewee Quotes from Sub-Theme 3.3 Radiographer Self-Efficacy

"Don't get me wrong I don't think they liked having to arbitrate, because it's quite a responsibility and I always used to say to them if you're not sure then just recall it"	*Director 2*
"I think we are all reasonably strong and we are all reasonably, we don't worry too much about saying no I disagree I wouldn't bring that back"	*Advanced Practitioner 1*
"I'd love to do it, but I'm not going to be allowed to because as I say they (Radiologists) think we would want to bring everything back"	*Advanced Practitioner 3*
"So, when I started third reading, I did them in tandem to start with. I would do them first, and then somebody else would like fourth read them almost after me. And I found that really valuable and reassuring"	*Consultant Radiographer 4*
"We've got a consultant who is about to start training in screening having got a lot of experience in symptomatic, and I can see that she might sort of say oh I don't want to do third reads and we'll be saying well do them and then you know, one of us will read it as well, and you check back and see what you think"	*Radiologist 1*
"There are advanced practitioners who don't want to do the 3rd reads I've heard them say I don't want that responsibility. Now I would talk to them about that if I thought they were good and experienced you know"	*Radiologist 1*
"I think it whether as film readers, it would be nice for us to arbitrate on our own. I don't know; I sit on the fence I think a little bit, I think because from a recent personal experience. Two radiographers were reporting together, and we missed, um, a cancer which came back a few weeks later, ok it was still picked up, but now I'm thinking perhaps, with us arbitrating, but then when we evaluated it all, it wasn't a massive barn door obvious one. It was one of these subtle, you look at and sort of talk yourself out of it, you know?	*Advanced Practitioner 2*
"It was always a bit that we would ask about or actually for a little while we kept our heads down about it because we just figured that if suddenly they decided we could do it, then we'd end up doing all of it (Laughter). So we kind of didn't shout too loudly about it to start with"	*Consultant Radiographer 4*

Table 47-3D Interviewee Quotes from Sub-Theme 3.4 Outcome expectancy and radiographer training

"I don't think we would consider them (radiographers) I think it would still go to an experienced radiologist because when I look at the quadrants the radiographer film readers, they do perform in a slightly different manner. Sometimes they tend to be to, to, they sit in two quadrants either very sensitive but not very specific, i.e. pick up cancers but call a huge amount to do that or else they tend to call a lot and miss a lot. So that tends to be the two quadrants they vary generally, but they tend to have recall rates that are probably twice that of all of the consultant"	*Director 4*
I think she (radiographer)probably does recall slightly more than the rest of us for third read but not to an extent where I'd you know be concerned about it"	*Radiologist 1*
"I think it's the way they've been trained; I think they are not used to making decisions ermm and having necessarily that amount of responsibility for their decisions. Hmm so they are more likely to air on the side of caution whereas doctors are inherently more comfortable with taking risks"	*Director 4*
"I think we're a bit more assertive, but in our particular practice, some of the film readers are a little bit reluctant to commit. Definitely, they find it; they struggle to make that final decision. Somebody has to make it, and we're a bit more decisive I would say the radiologists are a bit more decisive. I think that doctors in general in their training have a bit more of that. So, we have those sorts of qualities perhaps we have a bit more of those sort of qualities"	*Director 6*
"I think that's the problem with film readers who are not involved in assessment. They have a sort of training and education to a point, and then it stops"	*Director 3*

"They are a little bit uncertain of themselves, they may be a little bit more introvert or something like that a bit more anxious they don't want to make a mistake, and then there are other people who are just quite willing to say the first thing that comes into their head (laughter) you know and stick to it"	*Director 6*
"So, where I worked before, we had a lot more advanced practitioners making, doing reading and my perception is a lot of them wouldn't feel confident enough to override some, you know and say RR"	*Radiologist 1*
"You know consensus situation it's always been very much a discussion, but I would probably always bring them back, and again it would be a question of auditing and just seeing whether you're justified in bringing all of those back or not or what ones you didn't bring back"	*Consultant Radiographer 1*
"Because when I look at a film if it's something absolutely tiny I'll be thinking that is impossible to assess ermm and an assessment is not necessarily going to get the answer whereas somebody is just picking up every tiny little bit they can ermm yes it is a different decision-making process"	*Director 4*
"In order to be able to read against other radiographers, our clinical director said that he would like our sensitivity to be 90 percent or higher. So that was kind of, and that's an interpretation of whether you're a good film reader because experience and being good is two different things, isn't it? "	*Consultant Radiographer 4*

Table 47-3E Interviewee Quotes from Sub-Theme 3.5 Decision-making skills

"I want people who understand and have that accountability for making the decisions not just filtering through and going oh well that might be something I'm going to call that back, oh well I can't say it's not so I'll call it back and not having any concept of what you are going to put that woman through to prove it was nothing"	*Consultant Radiographer 2*
"Whereas the film reading radiographers who are radiographers during assessments I think were less specific because they didn't have that intensity of feeling about oh my god, I can't face assessing that (laughter) which is what it's about really"	*Director 1*
"I do think the point, third point about participate fully in assessment clinics including decision making, I think that's really important because I do think sometimes things are called and, you know, an advanced practitioner, will call them, but actually when you're the person actually doing the assessment, you do think oh goodness, how on earth am I going to you know, assess this, it's going to be, it's really, really tiny, you know, and I don't think if they don't have an appreciation of the assessment process, they're not the one actually doing it, I think it does make a difference. So, I think that's quite a valid point"	*Consultant Radiographer 1*
"With the locums is if you're not there all the time and you're not taking responsibility (so to speak), you tend to be more cautious about calling things normal. I think that does happen with people once they retire as well, they start to become much more cautious, and they just want to make sure that they don't miss anything, or be judged badly*"*	*Consultant Radiographer 2*
"If you had an educational sort of bench PERFORMS then I think that would be a much better way of improving people's reading skills"	*Director 1*

"So there's quite a few cases I've been reading since 2010 and been quite a few cases where I've said, I've been the one putting my foot down and saying no that just needs to come back"	*Advanced Practitioner 3*
"We are very much still leaning on the radiologist because we do the films in clinic, but we will probably help discuss sometimes. They'll ask us for our opinion, but they are ultimately, they are the ones with a patient in front of them and they're the ones having to put their name to it"	*Advanced Practitioner 2*
"That's the other problem with the arbitration and also again I think looking at it you know what the first and the second reader have said so people won't make their own independent judgement because they will be swayed"	*Director 3*
"I think they don't want to make the final decision. Absolutely. Definitely. And it really worries me because it means that the people who are doing arbitration and take to some extent (sorry I need to cough). People who are doing arbitration then have an ultimate responsibility which can be stressful, and secondly, we all have weaknesses in screen reading"	*Breast Clinician 1*
"The film readers are a little bit reluctant to commit definitely they find it, they struggle to make that final decision"	*Director 6*
"It's two things isn't it, one is picking up something that might be abnormal on the mammogram and the second thing is assessing its likelihood of being a cancer you know and obviously you're going to get it wrong and you are going to miss things you going to get it wrong occasionally but you want to keep those to a minimum number"	*Locum Radiologist 1*

Table 47-4A Interviewee Quotes from Sub-Theme 4.1 Collaborative working

"I think arbitration by consensus allows much more dissemination of good practice."	*Director 1*
"It's educational we often have some non-film reading radiographers in as well you know. As many as possible will attend that meeting for educational purposes so that they can see what they are looking at and it is so important"	*Advanced Practitioner 3*
"We wanted a time when all of us could actually have protected time to look at things and not just do the consensus but to do, to look at any interesting intervals or cases that have come up that week, so it seemed more sensible than one person on their own to have a look"	*Consultant Radiographer 3*
"I think it is a really good education tool when you can do it as a group; it's a really good opportunity to see what other people's perception is. I don't think it matters how experienced you are. I think it's; you can always pick up something new"	*Consultant radiographer 4*
"One person acts like Simon Cowell (laughter, and he takes the mouse, and he puts the report in, and we turn around and say oi we haven't made a decision, yet we haven't looked at them properly. Well, there's nothing there, get back to it. We have got a good enough team relationship to be able to say, go back, you're not just going to put normal/normal, we're going to see and ermm yeah"	*Advanced Practitioner 3*
"Everybody has an equal say really there's not a lead as such. So, no we don't really have a lead we kind of make a group decision"	*Director 5*
"We don't segregate radiologists at the front and radiographers at the back"	*Director 7*
"They weren't so worried about consensus meetings when that was written on the grounds that consensus meetings, there's some degree of, well at least its open people can see what's going on"	*Radiologist 3*
"The whole point is that we are all different aren't we. Some people are good at distortions, some people are good at calc, so you want to embrace all those different bits and have, and try and get the best, because it's a group isn't it and it only works as a collection of people looking at different things"	*Radiologist 2*

Table 47-4B Interviewee Quotes from Sub-Theme 4.2 Team dynamics and collegial conflict

"It's not a learning environment it's quite aggressive at times, I don't like it"	*Radiologist 2*
"Well sometimes you can say it's consensus, but it's a single individual that's making all of the decisions. Which, by definition, is not what it is, isn't it? I think it's very hard to make sure that doesn't happen which is quite difficult I think you have to make a real conscious effort to make sure that if you've got strong personalities that they're going to stay quiet"	*Director 3*
"Absolutely I think sometimes people are afraid to give their opinion especially if they've called something they are unwilling to back down and the strongest character will win"	*Director 4*
"At the moment somebody says that should definitely come back and if then they're aggressive about it then we all get a bit scared and say fine, whatever, I don't care"	*Radiologist 2*
"It becomes quite unpleasant with some people; it's kind of an attitude they've got because they will you know pull rank a lot of the time and say this is for me to decide. With you know some people just wanting to be top dog and, and put everybody else down"	*Advanced Practitioner 1*
"We do have a number of staff who are a lot quieter, and you know they do just bow down to the radiologist and keep their mouth shut and say nothing. People should be able to speak up"	*Advanced Practitioner 1*
"But it's a depressing way to start the day if it's just your opinion gets roughshod over and whoever, you know shouts the loudest gets their view across"	*Radiologist 2*
"I suggested that what we did was that we had a golden recall alarm a bit of a wacky idea (laughter), but basically if one reader really wants to recall it, they can use a golden recall, and they've only got twelve a year.	*Radiologist 3*

So, if they use a golden recall, it doesn't matter what anyone else says it's coming back end of story"	
"I will take a vote, and I have occasionally put the opinion of the meeting which isn't necessarily my own because I have been outvoted"	*Director 2*
"There are a lot of us as well; we have 7, 8, 10 film readers to fit in a small room and look at some pictures on one monitor, it's just not feasible. Well you can't have an opinion on something you can't see so you just sit there at the back going ooh I don't agree probably, but I can't see anyway"	*Radiologist 2*
"I think it reinforces increasing your recall rate which in a place that has a high recall rate, it doesn't help us at all; I don't think. it makes us read the same which is what I think our type of group arbitration does"	*Radiologist 2*
"I know that if I am fighting to get the you know, and they are saying oh no I don't want to see her I don't want to see her, well actually I do and I know that I can put her on my list"	*Consultant Radiographer 3*

Table 47-4C Interviewee Quotes from Sub-Theme 4.3 Accountability

"Knowing that it's not going to influence the recall rate necessarily means that they are quite happy to put things through for discussion for arbitration, and even if they are not at all sure it's going to be cancer and they just really want a bit of confidence"	*Director 1*
"*I will sometimes call things just to generate a discussion or I won't call things because they have, again to generate a discussion again"*	*Consultant Radiographer 1*
"Because a lot of people do recall just because they want another person's opinion. And they're just perhaps not completely confident enough to say no I think that's alright"	*Advanced Practitioner 1*

"So, I suppose it's like sharing the blame either way, isn't it in a group?"	*Consultant Radiographer 4*
"I think it will be much more difficult as a third, as a single reader because you are out there on your own"	*Consultant Radiographer 2*
"And whether you like it or not I think the public's perception is of still that a doctor is better trained, a better person and whether we like it or not, in reality, we are measured in different ways by the public. I would suspect a savvy lawyer in court will say, but one person did call it; one person did spot it"	*Consultant Radiographer 2*

Table 47-5A Interviewee Quotes from Sub-Theme 5.1 Guideline Factors

"I'm not fully versed with the document, but I know this is where they said that radiographers can arbitrate provided that they are involved in the assessment setting"	*Director 3*
"I'm not the unit director anymore I wasn't aware of that guidance"	*Radiologist 1*
"It's not very detailed is it, I mean it just says you can do what you want really"	*Radiologist 2*
"Well it's quite brief, isn't it? I think it must be one of the shortest, um, PHE breast screening documents I've ever seen"	*Consultant Radiographer 4*
"It doesn't say anything about performance though does it -it should do, it should be a no brainer that you get your best readers to do your arbitration or if that's how you do it as a single reader 3rd read it should be your best person and as a director you should know who your best reader is at that time. Obviously, it will change won't it as to how somebody is doing with their reading but in general it should be fairly easy to pick out the bestest one"	*Radiologist 2*
"If there are advanced practitioners undertaking arbitration to arbitrarily stop them from doing that because they're not a consultant practitioner well on the basis of no evidence at all seems ridiculous"	*Director 5*
"I think the only thing that I would query is whether there should be a little bit more specific criteria about your film reading"	*Consultant Radiographer 4*
"There probably is a better way of doing it, but we don't know what it is exactly, and that's why the guidance isn't so didactic as it might be, I suppose"	*Radiologist 3*

"I'd like us all to be doing the same thing. It makes it a bit less woolly then, and there's no, well I think if you're going to have guidelines make them as firm as you can"	*Radiologist 2*
"It is a bit vague, isn't it? I interpreted it as actually being an assessor in clinic rather than ermm our advanced practitioners…rather than being a sort of what's it called -responsible assessor"	*Director 5*
"This consultant practitioner level so does that mean only consultant practitioners would be able to, I have got it here participate fully in assessment clinics working to a consultant practitioner level"	*Advanced Practitioner 3*
"Its whether we then kind of encourage them to get accredited as well. But I feel a little bit ermm sheepish about saying that because I haven't done mine yet (laughter). I think there's definitely scope for it"	*Consultant Radiographer 4*
"Some aspects of it ermm almost feel like they are directed to exclude advanced practitioners doing single read arbitration, that justification that you have to work to a consultant practitioner level"	*Director 5*
"So, I thought oh right we've got to change our practice then because we are not following the arbitration guidance"	*Director 2*
"I'm not sure why it is there because the thing is if they've got the role as a film reader they don't need accreditation from the college for that because they've been, they've gained that role because they've got the experience and the trust has said that they are able to deliver on that role"	*Director 3*
"It is a controversial statement, and it's a bit, it may have been added inadvertently really. Maybe they didn't realise the significance to what they were putting in there. I don't know if anybody has questioned it or anything like that but ermm, but it does seem a bit odd I have to say"	*Director 6*

"Well I have not looked at that link so, so I don't quite know what that means"	*Director 2*
"It's a bit ambiguous really ermm accredited, I mean if somebody has an accreditation with their professional body I would then think that they you know to do advanced practice they should then be able, I should then be able to delegate you know doctor duties to them and they should perform them to the same standard"	*Director 4*
"Unless there's a tangible benefit then I'm not sure why you would need to do that because the ultimate responsibility lies with your organisation, because the college won't turn around and take responsibility"	*Director 3*
"Yes, I mean it's sort of like one of these waffly guidance things that nobody quite knows what to do with and causes a lot of heartache. So, it's either you are accredited to arbitrate, or it's not necessary because the fact that your screen reading to certain standards means you should be able to do all aspects of screen reading"	*Breast Clinician 1*

Table 47-5B Interviewee Quotes from Sub-Theme 5.2 Individual professional factors

"I think it's a little bit wishy-washy, you know, for the accreditation because it's because it's not mandatory like your HCP registration. You know, I think it's rather wishy-washy, yes, the society can provide accreditation, and it is on their website and everyone ignores it"	*Advanced practitioner 2*
"You know there always is; people will jealously guard their own little corner and don't wish to have other people prove themselves to be as good as them or better"	*Advanced practitioner 1*
"I think individual units have probably gone their own way and tried their own with that well before the advice came out"	*Locum Radiologist 1*
"I think it would be very interesting to know I don't feel this Public Health England guidance is very helpful I would be interested to know which radiologists and advanced practitioner's radiographers had been involved in setting up the guidance. My strong suspicion is people who don't really understand the way that we work has done that guidance"	*Radiologist 1*
"We thought it sort of applied 248 more to single arbitrators which we didn't do"	*Director 5*
"You see it says here which I thought was a strange thing to put in, if a new consultant radiologist then full appropriate training must have been completed. But you see I wouldn't have said that anybody who was new, to a role would be an appropriate person to be undertaking arbitration. I would have thought that it would be somebody who's got a lot of all-round experience because if they're new to the role they won't have"	*Advanced Practitioner 1*
"Nothing will change since the guidance has been produced"	*Breast Clinician 1*

"Basically, there has to be a radiologist taking the lead in consensus meetings we don't have consensus meetings which are just being conducted by film reading radiographers. So that's the way I look at it I assume that would meet those guidelines"	*Director 6*
"To be honest I didn't really take that much notice of it, I looked at it and just saw that we were, whatever we were doing was within those guidelines as in there is no radiographer arbitration"	*Director 4*
"well then that would mean that our advanced practitioner couldn't. I think then you're limiting yourself to a handful of radiographer consultants throughout the country. Well I think that statement is unhelpful, I would say, I think they need I think they need to attend MDT regularly and see the outcome of you know things that they've recalled so I would put more emphasis on MDT attendance"	*Radiologist 1*
"It was pretty much that I was employed in place of a consultant radiologist"	*Consultant Radiographer 2*
"Service need"	*Consultant Radiographer 4*
"It was causing unnecessary delays within the system if I wasn't arbitrating"	*Consultant Radiographer 2*
"When that came out in, yeah that's when we took it on board and we thought right okay, let's do this because we realised it was going to be of benefit to the department really, and the flow and it made sense, you know"	*Consultant Radiographer 1*
"To me, the criteria that they are using I think should be evidenced by whoever is doing this because I would suspect there are some units as well, and we are certainly, where radiologists are not meeting some of this criteria – you know"	*Consultant Radiographer 2*

"I think in some ways it has taken power; it's sort of implying that our, for us that our radiographer film readers some of whom are extremely good are not as good as radiologists. And I think that's an unfortunate message to have sent out"	*Director 2*
"We had to fight so hard for double radiographer reading. They held off on that for years and wouldn't let us do it. But all the research has shown now that you know that you're just as good at the reading. But because of staff shortages, we are doing double reading now and to be quite honest the radiographers are reading probably 3/4 of all screen reading because there's a few of us"	*Advanced practitioner 1*